

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 17:09:11 ; Search time 192 Seconds
(without alignments)
3001.422 Million cell updates/sec

Title: US-10-645-335-2

Perfect score: 7773

Sequence: 1 MAGNSLVLPVLWGRKAPTH.....TSNRNVILMAHDGKHFPMV 1490

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7773	100.0	1490	8	ADQ60212 Human rab
2	7772	100.0	1490	4	AAE10802 Human TGF
3	7763	99.9	1490	5	ABP70125 Human NOV
4	7418.5	95.4	1489	4	AAE10803 Mouse TGF
5	7373	94.9	1488	4	AAE10801 Rat TGF-b
6	3128.5	40.2	1525	4	ABB66447 Drosophil
7	3128.5	40.2	1525	8	ADS96554 Drosophil
8	2805.5	36.1	559	8	ADA55388 Human pro
9	1442	18.6	775	7	ADM04809 Human pro
10	1003	12.9	209	4	ADM19863 Protein e
11	909	11.7	210	4	ADM19833 Protein e
12	332	4.3	81	4	ADM20106 Protein e
13	228.5	2.9	579	8	ADS29860 Bacterial
14	228.5	2.9	608	8	ADS41794 Bacterial
15	220.5	2.8	1005	8	ADS41812 Bacterial
16	205.5	2.6	626	8	ADS41817 Bacterial
17	202	2.6	1136	8	ADS30013 Bacterial
18	199	2.6	1240	4	ABB61596 Drosophil
19	197.5	2.5	429	8	ADS22315 Bacterial
20	197	2.5	580	8	ADR86389 Aspergill
21	194.5	2.5	514	6	ABO07190 Human p53
22	191.5	2.5	478	8	ADS29872 Bacterial
23	187.5	2.4	459	4	ADM19860 Protein e
24	187.5	2.4	514	8	AB95225 Human pro
25	187.5	2.4	514	8	ADJ75513 Marker ge

Query Match 100.0%; Score 7773; DB 8; Length 1490;

26	186	2.4	573	3	AAB53475	AAB53475 Human col
27	185.5	2.4	679	8	ADS41629	ADS41629 Bacterial
28	185	2.4	972	4	AAM24010	AAM24010 Hamster E
29	185	2.4	972	4	AAM23988	AAM23988 Hamster E
30	182	2.3	943	6	ABR53214	ABR53214 Protein s
31	182	2.3	943	7	ADK63734	ADK63734 Disease t
32	181.5	2.3	514	8	ADJ76263	ADJ76263 Marker ge
33	178.5	2.3	1278	8	ABM80447	ABM80447 Tumour-as
34	178.5	2.3	1278	8	ADR96784	ADR96784 Human SRE
35	178	2.3	610	8	ADS41755	ADS41755 Bacterial
36	176.5	2.3	886	4	AAB93996	AAB93996 Human pro
37	174.5	2.2	579	8	ADM23920	ADM23920 Bacterial
38	174.5	2.2	587	2	AAV03204	AAV03204 Amino aci
39	173.5	2.2	1278	8	ADS10766	ADS10766 Human the
40	173	2.2	1481	7	ADC31541	ADC31541 Human nov
41	171	2.2	395	5	ABP51424	ABP51424 Human MDD
42	171	2.2	981	5	ABP73564	ABP73564 Candida a
43	169.5	2.2	358	4	AB965223	AB965223 Drosophil
44	169.5	2.2	1140	8	ADS30020	ADS30020 Bacterial
45	167.5	2.2	485	4	AAB68284	AAB68284 Amino aci

ALIGNMENTS

RESULT 1

ADQ60212
ID ADQ60212 standard; protein; 1490 AA.

XX ADQ60212;

XX 07-OCT-2004 (first entry)

XX Human rabconnectin 3-binding protein.

XX human; rabconnectin 3; GDP/GTP exchange reaction promotion protein;
binding protein; calcium dependent exocytosis.

XX Homo sapiens.

XX JP2004201673-A.

XX 22-JUL-2004.

XX 13-AUG-2003; 2003JP-00207500.

XX 01-NOV-2002; 2002JP-00319521.

XX (EISA) EISAI CO LTD.

XX WPI; 2004-556370/54.

XX N-PSDB; ADQ60211.

PT Novel protein that couples with rabconnectin 3 and GDP/GTP exchange
reaction promotion protein, useful for elucidating chemical substance
that activates or inhibits calcium dependent exocytosis.

PS Claim 1; SEQ ID NO 2; 52pp; Japanese.

CC The invention comprises the amino acid and coding sequence of a protein
that couples with rabconnectin 3 and GDP/GTP exchange reaction promotion
protein. The DNA and protein sequences of the invention are useful for
analysing immune tissue for locating the protein of the invention or for
analysing the expression level of the protein of the invention. The DNA
and protein sequences of the invention are useful for elucidating
chemical substances that activate or inhibit calcium dependent
exocytosis. The present amino acid sequence represents the human
rabconnectin 3 and GDP/GTP exchange reaction promotion protein-binding
protein of the invention.

SQ Sequence 1490 AA;

Best Local Similarity 100.0%; Pred. No. 0;			
Matches 1490; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MAGNSLVLPVLMGRKAPTHCISAVLLTDDGATVGTGCHDGOICLWDLSELVQLINPRALL	60
Db	1	MAGNSLVLPVLMGRKAPTHCISAVLLTDDGATVGTGCHDGOICLWDLSELVQLINPRALL	60
Qy	61	FHTASITCLSKACASSDKQYIIVSASESGEMCLWDVSDGRCEIETFKLACTHTGCIQYQPS	120
Db	61	FHTASITCLSKACASSDKQYIIVSASESGEMCLWDVSDGRCEIETFKLACTHTGCIQYQPS	120
Qy	121	VGNQREGRLCHGHYPILVVDATSLVLSYLSKISPDWISSMSIIRSHRTQEDTVAL	180
Db	121	VGNQREGRLCHGHYPILVVDATSLVLSYLSKISPDWISSMSIIRSHRTQEDTVAL	180
Qy	181	SVTGILKVMIVTSEISDMQDTEPIFEESKPIYCONCQISFCAFTQRSLLVVCVKYWRV	240
Db	181	SVTGILKVMIVTSEISDMQDTEPIFEESKPIYCONCQISFCAFTQRSLLVVCVKYWRV	240
Qy	241	FDAGDYSLLCSGSPSENGQTTGDFVSSDKVITWTENGQSYIYKLPASCLPASDSFRSDV	300
Db	241	FDAGDYSLLCSGSPSENGQTTGDFVSSDKVITWTENGQSYIYKLPASCLPASDSFRSDV	300
Qy	301	GKAVENLIPVQHILLDRDKKELLICPPVTRFPYGCREYFHKLLIQDSSGRLNIWNISD	360
Db	301	GKAVENLIPVQHILLDRDKKELLICPPVTRFPYGCREYFHKLLIQDSSGRLNIWNISD	360
Qy	361	TADQSGSEGLAMTTSISLQAEFDKLNPCPAGIIDQLSVIPNSNEPLKVTVASVYIPAHGR	420
Db	361	TADQSGSEGLAMTTSISLQAEFDKLNPCPAGIIDQLSVIPNSNEPLKVTVASVYIPAHGR	420
Qy	421	LVCGRDGSIVVPATQATVOLLQGEHMLRGWPHRTLRGRNKVTCLLYIYHOVSARY	480
Db	421	LVCGRDGSIVVPATQATVOLLQGEHMLRGWPHRTLRGRNKVTCLLYIYHOVSARY	480
Qy	481	DQRYLSGGVDVFSVIIWDIFSGEMKHIFCVHGGEITQLLVPENCARVOHCICSVASDH	540
Db	481	DQRYLSGGVDVFSVIIWDIFSGEMKHIFCVHGGEITQLLVPENCARVOHCICSVASDH	540
Qy	541	SVGLLSLREKCCIMLASRHLFPPIQVIKWRPDDYLTVGCSGDGVVYQMDTGALDRCVMG	600
Db	541	SVGLLSLREKCCIMLASRHLFPPIQVIKWRPDDYLTVGCSGDGVVYQMDTGALDRCVMG	600
Qy	601	ITAVEILNACDEAVPAADVSLSHPAVNLKQAMTRRSALAKNMAHKKLOTLATNLASEA	660
Db	601	ITAVEILNACDEAVPAADVSLSHPAVNLKQAMTRRSALAKNMAHKKLOTLATNLASEA	660
Qy	661	SDKGNLPKYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIQLLTEASRNTALISPENL	720
Db	661	SDKGNLPKYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIQLLTEASRNTALISPENL	720
Qy	721	QKASGSDKGSFLTGKRAAVLFQOVKETIKENIKEHLDDDEDEBEIMRQREESDPEY	780
Db	721	QKASGSDKGSFLTGKRAAVLFQOVKETIKENIKEHLDDDEDEBEIMRQREESDPEY	780
Qy	781	RSSKSRPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRLGMLKPHCTVSPGLLSR	840
Db	781	RSSKSRPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRLGMLKPHCTVSPGLLSR	840
Qy	841	GGHMSLMLPGYNOPACKLSHGKTEVGRKLPASGEVKGKTYGVSRVTTQHLLSIISLANT	900
Db	841	GGHMSLMLPGYNOPACKLSHGKTEVGRKLPASGEVKGKTYGVSRVTTQHLLSIISLANT	900
Qy	901	LMSMTNATFTGDHMKKGTPRPPRSTPDLISKARGPSPTSSNIIVQGOIKQVAAPVVSARS	960
Db	901	LMSMTNATFTGDHMKKGTPRPPRSTPDLISKARGPSPTSSNIIVQGOIKQVAAPVVSARS	960
Qy	961	ADHSIGSDPPSAPALHTCFLVNEGWSQLAMHCVMLPDLLGLDKFRPPLLEMLARRQDRC	1020
Db	961	ADHSIGSDPPSAPALHTCFLVNEGWSQLAMHCVMLPDLLGLDKFRPPLLEMLARRQDRC	1020
Qy	1021	LEVREAAQALLAELRRIEQAGKEAIDAWAPYLPQYIDHVISPGVTSEAAQTITTPADA	1080

Db	1021	LEVREAAQALLAELRRIEQAGKEAIDAWAPYLPQYIDHVISPGVTSEAAQTITTPADA	1080
Qy	1081	SGPEAKVQBEEDLVDDDITTTGCLSSVPQMKKISYSYERRKOATAIVLLGVIGAEFGAE	1140
Db	1081	SGPEAKVQBEEDLVDDDITTTGCLSSVPQMKKISYSYERRKOATAIVLLGVIGAEFGAE	1140
Qy	1141	IEPPKLLTPRSSSOIPEGFGLTSGGSNYSLARHTCKALTFLLOPPSPKLPHPHSTIRRT	1200
Db	1141	IEPPKLLTPRSSSOIPEGFGLTSGGSNYSLARHTCKALTFLLOPPSPKLPHPHSTIRRT	1200
Qy	1201	AIDLIGRGFTVWEPMYMDVSAVLMGLLELCADAQKOLANITMGLPLSPAADSARSARHALS	1260
Db	1201	AIDLIGRGFTVWEPMYMDVSAVLMGLLELCADAQKOLANITMGLPLSPAADSARSARHALS	1260
Qy	1261	LIATARPAPAFITIAKEVHRHTALAANTOSQONMMITTLARAKGILRLVIEILIEKMPD	1320
Db	1261	LIATARPAPAFITIAKEVHRHTALAANTOSQONMMITTLARAKGILRLVIEILIEKMPD	1320
Qy	1321	VVDLLVEVMDIIMYCLEGLSVKKGLQECFPAICRFYVSYERNHRIAVGARHGSVALY	1380
Db	1321	VVDLLVEVMDIIMYCLEGLSVKKGLQECFPAICRFYVSYERNHRIAVGARHGSVALY	1380
Qy	1381	DIRTKGQTIHGHKGPITAVAFAPDGRYLATYSNTDSDHISFWQNTSLGSGIMLSAPQ	1440
Db	1381	DIRTKGQTIHGHKGPITAVAFAPDGRYLATYSNTDSDHISFWQNTSLGSGIMLSAPQ	1440
Qy	1441	LRCIKTYQVPPVQPPASPSGSHNALKARLIWTSNRNVILMAHDGKEHFRMV	1490
Db	1441	LRCIKTYQVPPVQPPASPSGSHNALKARLIWTSNRNVILMAHDGKEHFRMV	1490
RESULT 2			
AAE10802			
ID	AAE10802 standard; protein; 1490 AA.		
XX	AAE10802;		
AC	AC		
DT	18-DEC-2001 (first entry)		
XX	Human TGF-beta Resistance Associated Gene (TRAG) protein.		
DE	Human; Transforming growth factor-beta; TGF-beta; cellular proliferation;		
KW	phenotype; TRAG; TGF-beta Resistance Associated Gene; cancer therapy;		
KW	oncogenesis; gene mapping; transgenic animal; chromosome 18.		
XX	Homo sapiens.		
OS	Homo sapiens.		
XX	Key		
FT	Domain	9..47	Location/Qualifiers
FT	Domain	/label= WD_repeat	
FT	Domain	56..95	
FT	Domain	/label= WD_repeat	
FT	Domain	239..246	
FT	Domain	/label= Tyrosine phosphorylation_motif	
FT	Domain	459..498	
FT	Domain	/label= WD_repeat	
FT	Domain	549..588	
FT	Domain	/label= WD_repeat	
FT	Domain	773..780	
FT	Domain	/label= Tyrosine phosphorylation_motif	
XX	WO200166739-A1.		
XX	13-SEP-2001.		
XX	12-FEB-2001; 2001WO-US004475.		
XX	07-MAR-2000; 2000US-0187572P.		
XX	08-MAR-2000; 2000US-0187848P.		
XX	(USSH) US DEPT HEALTH & HUMAN SERVICES.		
XX	Thorgerirsson SS, Sanders S;		

XX WPI; 2001-589942/66.
DR N-PSDB; AAD18130.
XX
PT Novel isolated polypeptide for studying and modulating mechanisms
PT involved in cellular proliferation comprises transforming growth factor-
PT beta resistance associated polypeptide fragment.
XX
XX Claim 1; Page 56-61; 89pp; English.
XX
CC The present sequence is human transforming growth factor (TGF)-beta
CC resistance associated (TRAG) protein. Human TRAG gene is located on
CC chromosome 18. TRAG DNA or protein is useful for studying and modulating
CC mechanisms involved in cellular proliferation, and for modulating
CC cellular phenotype. It is also used for screening aggressive metastasis
CC in cancer cells, and for targets used in cancer therapy. TRAG protein is
CC useful for evaluating factors that interact with and/or control TGF-beta
CC signalling for understanding both cell proliferation control and
CC oncogenesis, and in protein-protein interaction cells. TRAG DNA is useful
CC as hybridisation probe, in chromosome and gene mapping, in the generation
CC of antisense RNA and DNA, for the preparation of TRAG protein by
CC recombinant techniques, and for generating either transgenic or knockout
CC animals which, in turn, are useful in the development and screening of
CC therapeutically useful reagents
XX
SQ Sequence 1490 AA;
Query Match 100.0%; Score 7772; DB 4; Length 1490;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1489; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAGNSLVLPVLMGRKAPTHCISAVLLTDDGATVITGCHDQIQCLWDSVELQNPRL 60
DB 1 MAGNSLVLPVLMGRKAPTHCISAVLLTDDGATVITGCHDQIQCLWDSVELQNPRL 60
QY 61 FGTASTITCLSKACASDKQYIYVASESGEMCLWDYSDGRCIEFTKLACTHTGTGQYQPS 120
DB 61 FGTASTITCLSKACASDKQYIYVASESGEMCLWDYSDGRCIEFTKLACTHTGTGQYQPS 120
QY 121 VGNQREGRLCHGHYPVLLVVDATSLVLSYLSVKISPDWISSMSIIRSHRTQEDTVVAL 180
DB 121 VGNQREGRLCHGHYPVLLVVDATSLVLSYLSVKISPDWISSMSIIRSHRTQEDTVVAL 180
QY 181 SVTGILKWTIVTSEISDMQDTEPIFEESKPIYQNCQSIQSPCAFTQSLVVCVKYRW 240
DB 181 SVTGILKWTIVTSEISDMQDTEPIFEESKPIYQNCQSIQSPCAFTQSLVVCVKYRW 240
QY 241 FDAGDYSLLCGSPSENGQWTGSDYVSSDKVLIITWENGQSYIYKLPASCLPASDSFRSDV 300
DB 241 FDAGDYSLLCGSPSENGQWTGSDYVSSDKVLIITWENGQSYIYKLPASCLPASDSFRSDV 300
QY 301 GKAVENLIPVQHILLDRKDKELLIIPPVTRFPVGCYREYFKLLIOQDSSGRNLINWISD 360
DB 301 GKAVENLIPVQHILLDRKDKELLIIPPVTRFPVGCYREYFKLLIOQDSSGRNLINWISD 360
QY 361 TADKQSGEGLAMTTSISLQAEFDKLNPCPAGIIDQLSVIPNSNEPLKVYIIPAHGR 420
DB 361 TADKQSGEGLAMTTSISLQAEFDKLNPCPAGIIDQLSVIPNSNEPLKVYIIPAHGR 420
QY 421 LVCGREGDSIVIVPATQTAIVOLLQGEHMLRRCWPHRTLRGRNKVTCLLYPHQVSARY 480
DB 421 LVCGREGDSIVIVPATQTAIVOLLQGEHMLRRCWPHRTLRGRNKVTCLLYPHQVSARY 480
QY 481 DQRYLSGGVDFSVIIVIFSGEMKHIFCVHGGIEITQLLVPPENC SARVQHICSVASDH 540
DB 481 DQRYLSGGVDFSVIIVIFSGEMKHIFCVHGGIEITQLLVPPENC SARVQHICSVASDH 540
QY 541 SVGLLSLRKCKITMLASRLHFPPIQVIKWRPDDYLVVGCSDGSDYVYQMDTGALDRCVMG 600
DB 541 SVGLLSLRKCKITMLASRLHFPPIQVIKWRPDDYLVVGCSDGSDYVYQMDTGALDRCVMG 600
QY 601 ITAVEILNACDEAVPAVDLSLHPAVNLQAMTRRSALAKNWAHKLQTLATNLASEA 660
DB 601 ITAVEILNACDEAVPAVDLSLHPAVNLQAMTRRSALAKNWAHKLQTLATNLASEA 660

DB 601 ITAVEILNACDEAVPAVDLSLHPAVNLQAMTRRSALAKNWAHKLQTLATNLASEA 660
QY 661 SDKGNLPKYSHNSLWQAIKTNLTDPDIHVLFPDVEALIIQLLTERASRNTALISPENL 720
DB 661 SDKGNLPKYSHNSLWQAIKTNLTDPDIHVLFPDVEALIIQLLTERASRNTALISPENL 720
QY 721 QKASGSDKGGSLTGTCKRAAVLPOQVKETIKENIKEHLDDDEDEDEIMRQREESDPEY 780
DB 721 QKASGSDKGGSLTGTCKRAAVLPOQVKETIKENIKEHLDDDEDEDEIMRQREESDPEY 780
QY 781 RSSKSPKPLTLELVNLTWDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHCTVTSFGLLSR 840
DB 781 RSSKSPKPLTLELVNLTWDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHCTVTSFGLLSR 840
QY 841 GGHMSLMLPCYNQPACKLSHGKTEVGRKLPASGEGVGKGTGVSRVAVTQHLLSIISLANT 900
DB 841 GGHMSLMLPCYNQPACKLSHGKTEVGRKLPASGEGVGKGTGVSRVAVTQHLLSIISLANT 900
QY 901 LMSMTNATIGDHMKKGPTRPPSPDLSKARGSPPTSNNIVQGOIKQVAAFPVVSARS 960
DB 901 LMSMTNATIGDHMKKGPTRPPSPDLSKARGSPPTSNNIVQGOIKQVAAFPVVSARS 960
QY 961 ADHSGSDPPSAPALHTCFLVNEGWSOLAAMHCVMLPDLGLDKFRPPLLEMLARRWQDRC 1020
DB 961 ADHSGSDPPSAPALHTCFLVNEGWSOLAAMHCVMLPDLGLDKFRPPLLEMLARRWQDRC 1020
QY 1021 LEVREAAQALLAELRRIEQAGRKEAIDAWAPYLPQYIDHVISPGVTSEAAQTITAPDA 1080
DB 1021 LEVREAAQALLAELRRIEQAGRKEAIDAWAPYLPQYIDHVISPGVTSEAAQTITAPDA 1080
QY 1081 SGPEAKVQBEEDHLDVDDITTTGCLSSVPQMKISTSYEERRRKOATAIVLIGVIGABFGAE 1140
DB 1081 SGPEAKVQBEEDHLDVDDITTTGCLSSVPQMKISTSYEERRRKOATAIVLIGVIGABFGAE 1140
QY 1141 IEPPKLLTPRSSSQIPEGFGLTSGGSNYSLSRHTCKALTFLQLQPPSPKLPHPSTIRRT 1200
DB 1141 IEPPKLLTPRSSSQIPEGFGLTSGGSNYSLSRHTCKALTFLQLQPPSPKLPHPSTIRRT 1200
QY 1201 AIDLIGRGFTVMEPYMDVSAVLMGLLELCEADAEKQLANITMGLPLSPAADSARSARHLS 1260
DB 1201 AIDLIGRGFTVMEPYMDVSAVLMGLLELCEADAEKQLANITMGLPLSPAADSARSARHLS 1260
QY 1261 LIATARPAPFITIAKEVHRHTALAANTOSQONMHTTTLARAKEGLRVLRIEILKEMPTD 1320
DB 1261 LIATARPAPFITIAKEVHRHTALAANTOSQONMHTTTLARAKEGLRVLRIEILKEMPTD 1320
QY 1321 VDILLVEVMDIIMYCLEGSLVKKKGLOECFPAICRFYMYVYERNHRIAVGARHGSVALY 1380
DB 1321 VDILLVEVMDIIMYCLEGSLVKKKGLOECFPAICRFYMYVYERNHRIAVGARHGSVALY 1380
QY 1381 DIETGKQTIHGHKGPITAVAFAPDGRYLATYSNTDHSISFWQMTSLGSGIGMLNSAPO 1440
DB 1381 DIETGKQTIHGHKGPITAVAFAPDGRYLATYSNTDHSISFWQMTSLGSGIGMLNSAPO 1440
QY 1441 LRCIKTYQVPPVQSPGSHNALKARLIWTSNRNVILMAHDGKEHFRFW 1490
DB 1441 LRCIKTYQVPPVQSPGSHNALKARLIWTSNRNVILMAHDGKEHFRFW 1490
RESULT 3
ABP70125
ID ABP70125 standard; protein; 1490 AA.
XX
AC ABP70125;
XX
DT 27-JAN-2003 (first entry)
XX
DE Human NOV36a.
XX
KW Human; anti-HIV; cytostatic; antidiabetic; antiaesthetic; cachexia; AIDS;
KW antinflammatory; cardiant; haemostatic; neuroprotective; anorectic;
KW nootropic; immunosuppressive; osteopathic; antiparkinsonian; cancer;
KW antifertility; cerebroprotective; gene therapy; NOVX; NOV; fertility;

KW metabolic disorder; diabetes; obesity; infectious disease; anorexia;
 KW neurodegenerative disease; Alzheimer's disease; Parkinson's disease;
 KW immune disorder; haematopoietic disorder; cardiovascular disorder;
 KW bronchial asthma; dyslipidemia; metabolic disturbance; neurogenesis;
 KW metabolic syndrome X; wasting disorder; cell differentiation;
 KW cell proliferation; haematopoiesis; wound healing; angiogenesis.
 XX

OS Homo sapiens.

XX WO200272771-A2.

XX 19-SEP-2002.

PF 08-MAR-2002; 2002WO-US007288.

XX 08-MAR-2001; 2001US-0274101P.

PR 08-MAR-2001; 2001US-0274194P.

PR 08-MAR-2001; 2001US-0274281P.

PR 08-MAR-2001; 2001US-0274322P.

PR 09-MAR-2001; 2001US-0274849P.

PR 12-MAR-2001; 2001US-0275235P.

PR 13-MAR-2001; 2001US-0275578P.

PR 13-MAR-2001; 2001US-0275579P.

PR 14-MAR-2001; 2001US-0275601P.

PR 14-MAR-2001; 2001US-0276000P.

PR 16-MAR-2001; 2001US-0276766P.

PR 19-MAR-2001; 2001US-0276994P.

PR 20-MAR-2001; 2001US-0277239P.

PR 20-MAR-2001; 2001US-0277321P.

PR 20-MAR-2001; 2001US-0277327P.

PR 20-MAR-2001; 2001US-0277338P.

PR 21-MAR-2001; 2001US-0277791P.

PR 22-MAR-2001; 2001US-0277833P.

PR 23-MAR-2001; 2001US-0278152P.

PR 26-MAR-2001; 2001US-0278894P.

PR 27-MAR-2001; 2001US-0278999P.

PR 27-MAR-2001; 2001US-0279036P.

PR 28-MAR-2001; 2001US-0279344P.

PR 30-MAR-2001; 2001US-0279995P.

PR 30-MAR-2001; 2001US-0280233P.

PR 02-APR-2001; 2001US-0280802P.

PR 02-APR-2001; 2001US-0280822P.

PR 02-APR-2001; 2001US-0280900P.

PR 04-APR-2001; 2001US-0281194P.

PR 13-APR-2001; 2001US-0283675P.

PR 30-APR-2001; 2001US-0287424P.

PR 02-MAY-2001; 2001US-0288066P.

PR 03-MAY-2001; 2001US-0288342P.

PR 13-MAY-2001; 2001US-0291190P.

PR 16-MAY-2001; 2001US-0291099P.

PR 16-MAY-2001; 2001US-0291240P.

PR 30-MAY-2001; 2001US-0294485P.

PR 31-MAY-2001; 2001US-0294889P.

PR 31-MAY-2001; 2001US-0294899P.

PR 18-JUN-2001; 2001US-0299027P.

PR 19-JUN-2001; 2001US-0299303P.

PR 19-JUN-2001; 2001US-0299310P.

PR 10-JUL-2001; 2001US-0304354P.

PR 31-JUL-2001; 2001US-0309198P.

PR 16-AUG-2001; 2001US-0312903P.

PR 10-SEP-2001; 2001US-0318462P.

PR 12-SEP-2001; 2001US-0318770P.

PR 27-SEP-2001; 2001US-0325430P.

PR 27-SEP-2001; 2001US-0325681P.

PR 18-OCT-2001; 2001US-0330380P.

PR 31-OCT-2001; 2001US-0335301P.

PR 14-NOV-2001; 2001US-0332172P.

PR 14-NOV-2001; 2001US-0332271P.

PR 14-NOV-2001; 2001US-0332272P.

PR 14-NOV-2001; 2001US-0333184P.

PR 14-NOV-2001; 2001US-0333272P.

PR 21-NOV-2001; 2001US-0332094P.

PR 03-DEC-2001; 2001US-0337426P.
 PR 03-DEC-2001; 2001US-0338092P.
 PR 04-DEC-2001; 2001US-0337185P.
 PR 03-JAN-2002; 2002US-0345705P.
 PR 08-MAR-2002; 2002US-00093463.

XX (CURA-) CURAGEN CORP.

XX Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ;

PI Boldog FL, Li L, Zerhusen BD, Tchernev VT, Gangolli EA, Vernet CAM;

PI Pena CE, Burgess CE, Liu X, Spytek KA, Gorman L, Spaderna SK;

PI Voss EZ, Malyankar UM, Anderson DW, Patturajan M, Miller CE;

PI Taupier RJ, Padigar M, Shenoy SG, Kekuda R, Gusev VY, Pochart PF;

PI Zhong M;

XX WPI; 2002-732824/79.

DR N-PSDB; ABV99403.

XX New NOVX polypeptides and polynucleotides, useful for preventing,

PT diagnosing or treating NOVX-associated disorders e.g. diabetes, cancer,

PT Alzheimer's disease, dyslipidemia, obesity, immune or hematopoietic

PT disorders, and asthma.

XX Claim 1; Page 232; 619pp; English.

XX The present invention relates to new isolated proteins (NOVX) and their

CC coding sequences (ABV99327-ABV99595 and ABP70049-ABP70149), where X is

CC any number from 1 to 48. The NOVX proteins and coding sequences are

CC useful in the manufacture of a medicament for treating a syndrome

CC associated with a human disease, preferably a NOVX-associated disorder.

CC The NOVX coding sequences and proteins are useful for treating,

CC preventing or diagnosing diseases such as metabolic disorders, diabetes,

CC obesity, infectious disease, anorexia, cancer-associated cachexia,

CC cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's

CC disease, immune disorders, haematopoietic disorders, cardiovascular

CC disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic

CC disturbances associated with obesity, metabolic syndrome X or wasting

CC disorders associated with chronic diseases or various cancers. The NOVX

CC coding sequences and proteins may also be used as targets for the

CC identification of small molecules that modulate or inhibit e.g.

CC neurogenesis, cell differentiation, cell proliferation, haematopoiesis,

CC wound healing and angiogenesis, in gene therapy, in generation of

CC antibodies that bind immunospecifically to NOVX substances for use in

CC therapeutic or diagnostic methods

XX Sequence 1490 AA;

SQ

Query Match 99.9%; Score 7763; DB 5; Length 1490;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1488; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAGNSLVLPVILWGRKAPTHCISAVLLTDDGATVTCGHDGQICLDVLSVELQINPRALL 60
 Db 1 MAGNSLVLPVILWGRKAPTHCISAVLLTDDGATVTCGHDGQICLDVLSVELQINPRALL 60

QY 61 FGHASITCLSKACASSDKQYIVSASESGEMCLWDVSDGRCIEFTKLACTHTGIIQYQFS 120

Db 61 FGHASITCLSKACASSDKQYIVSASESGEMCLWDVSDGRCIEFTKLACTHTGIIQYQFS 120

QY 121 VGNQREGRLCHGHYPEILVVDATSLVLYSLVSKISPDWISMSIIRSHRTOEDTVVAL 180

Db 121 VGNQREGRLCHGHYPEILVVDATSLVLYSLVSKISPDWISMSIIRSHRTOEDTVVAL 180

QY 181 SVTGILKWIIVTSEISDMQDTEPIFEESKPIYQNCQCSISCAFTQRIILLVCSKYWRV 240

Db 181 SVTGILKWIIVTSEISDMQDTEPIFEESKPIYQNCQCSISCAFTQRIILLVCSKYWRV 240

QY 241 FDAGDYSLLCSGSENGQWTGDFVSSDKVIWTENGQSYIYKLPASCLPASDSFRSDV 300

Db 241 FDAGDYSLLCSGSENGQWTGDFVSSDKVIWTENGQSYIYKLPASCLPASDSFRSDV 300

QY 301 GKAVENLIPPVQHILLDRKDKELLICPPVTRFPYGCREFYHKLIIQGDSSGRININISD 360

Db 301 GKAVENLIPPVQHILLDRKDKELLICPPVTRFPYGCREFYHKLIIQGDSSGRININISD 360

Db 301 GKAVENLIPVQHILLDRDKKELLICPPVTRFFVGCBEYFHKLLIQDSSGRNLWNISD 360
Qy 361 TADKQSEEGLAMTTISLQAFDKLNPFCAGIIDLQSVIPNSNEPLKVTASVYPAHGR 420
Db 361 TADKQSEEGLAMTTISLQAFDKLNPFCAGIIDLQSVIPNSNEPLKVTASVYPAHGR 420
Qy 421 LVCGRDGSIVIPATQTAIVOLLOGEHMLRGHPHRTLRGHRNKVTCLLYPHQVSARY 480
Db 421 LVCGRDGSIVIPATQTAIVOLLOGEHMLRGHPHRTLRGHRNKVTCLLYPHQVSARY 480
Qy 481 DQRLISGGVDFSVIWDIFSGEMKHIFCVHGGBITQLVPPENC SARVOHCICSVASDH 540
Db 481 DQRLISGGVDFSVIWDIFSGEMKHIFCVHGGBITQLVPPENC SARVOHCICSVASDH 540
Qy 541 SVGLLSLREKCKIMLASRHLFPPIQVIKWRPSDDLVVVGCSGDSYVWMQDGTALDRCVMG 600
Db 541 SVGLLSLREKCKIMLASRHLFPPIQVIKWRPSDDLVVVGCSGDSYVWMQDGTALDRCVMG 600
Qy 601 ITAVEILNACDEAVPAAVDSLHPAVNLKQAMTRRSALAKNMAHHKLOTATNLLASEA 660
Db 601 ITAVEILNACDEAVPAAVDSLHPAVNLKQAMTRRSALAKNMAHHKLOTATNLLASEA 660
Qy 661 SDKGNLPKYSHNSLWQAIKTNLTDPDHLVFPDVEALIIQLLTERASRPNLTALISPENL 720
Db 661 SDKGNLPKYSHNSLWQAIKTNLTDPDHLVFPDVEALIIQLLTERASRPNLTALISPENL 720
Qy 721 QKAGSSDKGSGFLTGKRAAVLFOQVKETIKENIKEHLDDDEDEBEIMQRREESDPEY 780
Db 721 QKAGSSDKGSGFLTGKRAAVLFOQVKETIKENIKEHLDDDEDEBEIMQRREESDPEY 780
Qy 781 RSSKSKPLTLLVNLWDTAKLFMSCLHAWGLNEVLDEVCLDRLGMLKPHCTVTSFGLLSR 840
Db 781 RSSKSKPLTLLVNLWDTAKLFMSCLHAWGLNEVLDEVCLDRLGMLKPHCTVTSFGLLSR 840
Qy 841 GGHMSLMLPGYNOPACKLSHGKTEVGRKLPASEGVGKGTGVSRVAVTQHLSTIISLANT 900
Db 841 GGHMSLMLPGYNOPACKLSHGKTEVGRKLPASEGVGKGTGVSRVAVTQHLSTIISLANT 900
Qy 901 LMSMTNATFGDHMKGTPRPPSTPDLKARGSPPTSSNIYVQGIKQVAAAPVVSARS 960
Db 901 LMSMTNATFGDHMKGTPRPPSTPDLKARGSPPTSSNIYVQGIKQVAAAPVVSARS 960
Qy 961 ADHSGSDPPSAPALHTCFLVNEGWSQLAAHCVMLPDLLGLDKFRPPLLEMLARRWQDR 1020
Db 961 ADHSGSDPPSAPALHTCFLVNEGWSQLAAHCVMLPDLLGLDKFRPPLLEMLARRWQDR 1020
Qy 1021 LEVREAAQALLAELEIEQGRKEADAWAPYLPQVIDHVISPGVTSEAAQTITAPDA 1080
Db 1021 LEVREAAQALLAELEIEQGRKEADAWAPYLPQVIDHVISPGVTSEAAQTITAPDA 1080
Qy 1081 SGPEAKVQBEHDLVDDDIITGCLSSVPQMKKISTSYEERRKQATVLLGVIGABFGAE 1140
Db 1081 SGPEAKVQBEHDLVDDDIITGCLSSVPQMKKISTSYEERRKQATVLLGVIGABFGAE 1140
Qy 1141 IEPPKLLTRPRSSQIPEGFGLTSGGNSYLARHTCKALFTLLQPPSPKLPHPSTIRRT 1200
Db 1141 IEPPKLLTRPRSSQIPEGFGLTSGGNSYLARHTCKALFTLLQPPSPKLPHPSTIRRT 1200
Qy 1201 AIDLIGGFTWPEYMDVSAVLMGLLELCAEQLANITWGLPLSPAADSARSARHLS 1260
Db 1201 AIDLIGGFTWPEYMDVSAVLMGLLELCAEQLANITWGLPLSPAADSARSARHLS 1260
Qy 1261 LIATARPPAFTITAKEVHRHTALAANTQSQNNHTTTLARAKEILRVIEILIEKMPD 1320
Db 1261 LIATARPPAFTITAKEVHRHTALAANTQSQNNHTTTLARAKEILRVIEILIEKMPD 1320
Qy 1321 VDLLEWVMDIMYCLEGSLVKKKGLQECFPAICRFYVWSYERNHRIAVGARHGSVALY 1380
Db 1321 VDLLEWVMDIMYCLEGSLVKKKGLQECFPAICRFYVWSYERNHRIAVGARHGSVALY 1380
Qy 1381 DIRTGKQTHGHKGPITAVAPADGRYLATYNTDISHIFWQNTSLIGSIGMLNSAPQ 1440
Db 1381 DIRTGKQTHGHKGPITAVAPADGRYLATYNTDISHIFWQNTSLIGSIGMLNSAPQ 1440

Qy 1441 LRCIKTYQVPPVQSPASGSHNALKARLIWTSNRNVILMAHDGKEHRFMV 1490
Db 1441 LRCIKTYQVPPVQSPASGSHNALKARLIWTSNRNVILMAHDGKEHRFMV 1490
RESULT 4
AAE10803
ID AAE10803 standard; protein; 1489 AA.
XX AAE10803;
XX 18-DEC-2001 (first entry)
XX Mouse TGF-beta Resistance Associated Gene (TRAG) protein.
XX Mouse; Transforming growth factor-beta; TGF-beta; cellular proliferation;
XX phenotype; TRAG; TGF-beta Resistance Associated Gene; cancer therapy;
XX oncogenesis; gene mapping; transgenic animal; chromosome 18.
XX Mus sp.
XX
XX Key Location/Qualifiers
XX Domain 9..47
XX FT /label= WD_repeat
XX Domain 56..95
XX FT /label= WD_repeat
XX Domain 239..246
XX FT /label= Tyrosine phosphorylation_motif
XX Domain 459..498
XX FT /label= WD Repeat
XX Domain 549..588
XX FT /label= WD_repeat
XX Domain 773..780
XX FT /label= Tyrosine phosphorylation_motif
XX
XX WO200166739-A1.
XX
XX 13-SEP-2001.
XX
XX 12-FEB-2001; 2001WO-US004475.
XX
XX 07-MAR-2000; 2000US-0187572P.
XX 08-MAR-2000; 2000US-0187848P.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Thorgeirsson SS, Sanders S;
XX
XX WPI; 2001-589942/66.
XX N-PSDB; AAD18131.
XX
XX Novel isolated polypeptide for studying and modulating mechanisms
XX involved in cellular proliferation comprises transforming growth factor-
XX beta resistance associated polypeptide fragment.
XX
XX Claim 1; Page 62-66; 89pp; English.
XX
XX The present sequence is mouse transforming growth factor (TGF)-beta
XX resistance associated (TRAG) protein. Mouse TRAG gene is located on
XX chromosome 18. TRAG DNA or protein is useful for studying and modulating
XX mechanisms involved in cellular proliferation, and for modulating
XX cellular phenotype. It is also used for screening aggressive metastasis
XX in cancer cells, and for targets used in cancer therapy. TRAG protein is
XX useful for evaluating factors that interact with and/or control TGF-beta
XX signalling for understanding both cell proliferation control and
XX oncogenesis, and in protein-protein interaction cells. TRAG DNA is useful
XX as hybridisation probe, in chromosome and gene mapping, in the generation
XX of antisense RNA and DNA, for the preparation of TRAG protein by
XX recombinant techniques, and for generating either transgenic or knockout
XX animals which, in turn, are useful in the development and screening of
XX therapeutically useful reagents
XX

SQ	Sequence 1489 AA;	
Query Match	95.4%; Score 7418.5; DB 4; Length 1489;	
Best Local Similarity	95.0%; Pred. No. 0;	
Matches 1415; Conservative	36; Mismatches 38; Indels 1; Gaps 1;	
QY	1 MAGNSVLPTVLMGRKAPHCISAVLLTDDGATVITGCHDQICLMDLSVELQINPRALL 60	
DB	1 MAGNSVLPTVLMGRKAPHCISAVLLTDDGATVITGCHDQICLMDVSVLEVNPRALL 60	
QY	61 FGTASITCLSKACASDKQYIYVSASESGEMCLWDVSDGRCIBFTKLACTHTGTGQYQFS 120	
DB	61 FGTASITCLSKACASDKRYTYVSANGEMCLWDVNDGRCIBFTKLACTHTGTGQYQFS 120	
QY	121 VGNQREGRLCHGHYPEILVVDATSLVLYSLVSKISPDWISSMSIIRSHRTQEDTVVAL 180	
DB	121 VGNQREGRLCHGHYPEILVVDATSLVLYSLVSKISPDWISSMSIIRSHRTQEDTVVAL 180	
QY	181 SVTGILKVMITVTSISDMQDTEPIFEESKPIYCONCQISPCAFRTORSLLVVCYSKYRW 240	
DB	181 SVTGILKVMITVTSISDMQDTEPIFEESKPIYCONCQISPCAFRTORSLLVVCYSKYRW 240	
QY	241 FDAGDYSLLCSGFSENGQWTGDFVSSDKVIIWTENGQSYIYKLPASCLPASDSFRSDV 300	
DB	241 FDAGDYSLLCSGFSENGQWTGDFVSDKVIWTENGQSYIYKLPASCLPASDSFRSDV 300	
QY	301 GKAVENLIIPVQHILLDRDKKELLICPPVTRFPYGCREYFHKLLIQDSSGRINIWNISD 360	
DB	301 GKAVENLIIPVQHSHLLDQDKELVCPVTRFPYGCYKELHKLIIQDSSGRINIWNISD 360	
QY	361 TADQSGSEGLAMTTSISLQEAQDKLPCPAGIIDOLSVIPNSNEPLKVTASVYIPAHGR 420	
DB	361 IAEKQEADEGLKMTTCSISLQEAQDKLPCPAGIIDOLSVIPNSNEPLKVTASVYIPAHGR 420	
QY	421 LVCGREDGSIIVPATQTAIVQLLQGEHMLRRGPPHRTLGRHNKVTCLLYPHQVSARY 480	
DB	421 LVCGREDGSIIVPATQTAIVQLLQGEHMLRRGPPHRTLGRHNKVTCLLYPHQVSARY 480	
QY	481 DQYILSGGVDFSVIWDIFSGEMKHIFCVHGGEITQLLVPPENCARSARVQHICISVASDH 540	
DB	481 DQYILSGGVDFSVIWDIFSGEMKHIFCVHGGEITQLLVPPENCARSARVQHICISVASDH 540	
QY	541 SVGLLSUREKCCIMLASRHLFPQVIKWRPSDDYLVVCGSDGVVYVQMDTGALDCVMG 600	
DB	541 SVGLLSUREKCCIMLASRHLFPQVIKWRPSDDYLVVCGSDGVVYVQMDTGALDCVMG 600	
QY	601 ITAVEILNACDEAVPAVDLSLHPAVNLKQAMTFRSLAALKNNMAHKLQTLATNLLASEA 660	
DB	601 ITAVEILNACDEAVPAVDLSLHPAVNLKQAMTFRSLAALKNNMAHKLQTLATNLLASEA 660	
QY	661 SDKGNLPKYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIQLLTEASRPNTALISPENL 720	
DB	661 SDKGNLPKYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIQLLTEASRPNTALISPENL 720	
QY	721 QKASGSDKGSFLLTGKRAAVLFQOVKETIKENIKHLLDDEDEBEIMRQREESDPEY 780	
DB	721 QKASGSDKGSFLLTGKRAAVLFQOVKETIKENIKHLLDDEDEBEIMRQREESDPEY 780	
QY	781 RSSKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDLGLMKPHCTVSGLLSR 840	
DB	781 RASKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDLGLMKPHCTVSGLLSR 840	
QY	841 GGHMSLMLPGYNQPAACKLSHGKTEVGRKLPASGCVGKGTGVSRVATTQHLLSIISLANT 900	
DB	841 GGHMSLMLPGYNAAGKLLHAKAEVGRKLPAAEGVGKGTGVSRVATTQHLLSIISLANT 900	
QY	901 LMSMTNATFGDMMKGPTRPPPTPDLISKARGSPPTSSNIYQGOIKOVAAPVWSARSD 960	
DB	901 LMSMTNATFGDMMKGPTRPPPTPDLISKARDSPPASSNIYQGOIKQAAAAPVWSARSD 960	
QY	961 ADHSGSDPPSAPALHTCFLVNEGWSQLAAMHCVMLPDLGLDKFRPPLLEMLARRWQDRC 1020	
DB	961 ADHSGSDSAS - PALPTCFLVNEGWSQLAAMHCVMLPDLGLGLERFRPPLLEMLARRWQDRC 1019	

QY	1021 LEVRAAAQALLAEELRRIEQAGRKBAIDAWAPVLPOYIDHVISPGVTSEAAOTITTAPDA 1080	
DB	1020 LEVRAAAQALLAEELRRIEQAGRKETIDTWAPYLPQYMDHVISPGVTAGAMQMAAPDA 1079	
QY	1081 SGPEAKVQSEEHDLVDDDDITTCGLSSVPQMKKISTSYERRRKOATAIVLLGVIGAEFGAE 1140	
DB	1080 SGPEAKVQSEEHDLVDDDDITAGCLSSVPQMKKISTSYERRRKOATAIVLLGVIGAEFGAE 1139	
QY	1141 IEPKLLTRPRSSQIPGFGLTSGGSNYSLARHTCKALTFLLLQPPSPKLPHPHSTIRRT 1200	
DB	1140 IEPKLLTRPRSSQIPGFGLTSGGSNYSLARHTCKALTYLLQPPSPKLPHPHSTIRRT 1199	
QY	1201 ATDLTGRGTWPEPYMDVSAVLMGLLELCADAEKOLANITMGLPLSPAADSARSARHALS 1260	
DB	1200 ATDLTGRGTWPEPYMDVSAVLMGLLELCADAEKOLANITMGLPLSPAADSARSARHALS 1259	
QY	1261 LIATARPAPAFITTIAREVHRHTALAANTOSQOMHTTTTLARAKGEILRVIIELEKMPDT 1320	
DB	1260 LIATARPAPAFITTIAREVHRHTALAANTOSQSIHTTLARAKGEILRVIIELEKMPDT 1319	
QY	1321 VVDLLVEVMDIIMYCLEGLSVKKGLOECFPAICRFYMWVSYTERNHRITAVGARHGSVALY 1380	
DB	1320 VVDLLVEVMDIIMYCLEGLSVKKGLOECFPAICRFYMWVSYTERSHRIAVGARHGSVALY 1379	
QY	1381 DIRTGCKQTIHGKGPITAVAFAPDGRVLYATYSNTDSDHISFWQMTSLLSIGMLNSAPQ 1440	
DB	1380 DIRTGCKQTIHGKGPITAVFAPDGRVLYATYSNTDSDHISFWQMTSLLSIGMLNSAPQ 1439	
QY	1441 LRCIKTYQVPPVQSPASPGSHNALKLARLIWTSNRNVILMAHDGKEHRFMV 1490	
DB	1440 LACIKTYQVPPVQSPASPGSHNALKLARLIWTSNRNVILMAHDGKEHRFMV 1489	
RESULT 5		
AAE10801	ID	AAE10801 standard; protein; 1488 AA.
XX	AC	AAE10801;
XX	XX	
DT	18-DEC-2001	(first entry)
XX		
DE		Rat TGF-beta Resistance Associated Gene (TRAG) protein.
XX		
KW		Rat; Transforming growth factor-beta; TGF-beta; cellular proliferation;
KW		phenotype; TRAG; TGF-beta Resistance Associated Gene; cancer therapy;
KW		oncogenesis; gene mapping; transgenic animal.
XX		
OS		Rattus sp.
XX		
FT	Key	Location/Qualifiers
FT	Domain	9..47
FT		/label= WD_repeat
FT	Domain	56..95
FT		/label= WD_repeat
FT	Domain	239..246
FT		/label= Tyrosine_phosphorylation_motif
FT	Domain	459..498
FT		/label= WD_repeat
FT	Domain	549..588
FT		/label= WD_repeat
FT	Domain	773..780
FT		/label= Tyrosine_phosphorylation_motif
XX		
PN		WO200166739-A1.
XX		
PD		13-SEP-2001.
XX		
PF		12-FEB-2001; 2001WO-US004475.
XX		
PR		07-MAR-2000; 2000US-0187572P.
PR		08-MAR-2000; 2000US-0187848P.
XX		

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Thorgeirsson SS, Sanders S;

WPI; 2001-589942/66.

N-PSDB; AAD18129.

Novel isolated polypeptide for studying and modulating mechanisms involved in cellular proliferation comprises transforming growth factor-beta resistance associated polypeptide fragment.

Claim 1; Page 51-56; 89pp; English.

The present sequence is rat transforming growth factor (TGF)-beta resistance associated (TRAG) protein. TRAG DNA or protein is useful for studying and modulating mechanisms involved in cellular proliferation, and for modulating cellular phenotype. It is also used for screening aggressive metastasis in cancer cells, and for targets used in cancer therapy. TRAG protein is useful for evaluating factors that interact with and/or control TGF-beta signalling for understanding both cell proliferation control and oncogenesis, and in protein-protein interaction cells. TRAG DNA is useful as hybridisation probe, in chromosome and gene mapping, in the generation of antisense RNA and DNA, for the preparation of TRAG protein by recombinant techniques, and for generating either transgenic or knockout animals which, in turn, are useful in the development and screening of therapeutically useful reagents

XX SQ Sequence 1488 AA;

Query Match 94.9%; Score 7373; DB 4; Length 1488;

Best Local Similarity 94.5%; Pred. No. 0;

Matches 1408; Conservative 39; Mismatches 41; Indels 2; Gaps 2;

QY 1 MAGNSLVLPVLWGRKAPTHCISAVLLTDDGATVITGCHDGOICLWLSVELQINPRALL 60
DB 1 MAGNSLVLPVLWGRKAPTHCISAVLLTDDGATVITGCHDGOICLWLSVELQINPRALL 60
QY 61 FGHASITCLSKACASDQYIVSASEGEMCLWVSDGRCIBFTKLACTHTGIGYQFS 120
DB 61 FGHASITCLSKACASDQYIVSASEGEMCLWVSDGRCIBFTKLACTHTGIGYQFS 120
QY 121 VGNQREGRLCHGHYPEILVVDATSLVLSYLSKISPDWISSMIIIRSHRTQEDTVVAL 180
DB 121 VGNQREGRLCHGHYPEILVVDATSLVLSYLSKISPDWISSMIIIRSHRTQEDTVVAL 180
QY 181 SVTGILKWLVIIVTSEISDMQDTEPIFEESKPIYQNCQSFCAFTQSLVVCYSKYRV 240
DB 181 SVTGILKWLVIIVTSEISGLQDTEPIFEESKPIYQNCQSFCAFTQSLVVCYSKYRV 240
QY 241 FDAGDYSLLCSGSPENGQWTGDFVSSDKVIIWTENGQSYIYKLPASCLPASDSFRSDV 300
DB 241 FDAGDYSLLCSGSPENGQWTGDFVSDKVIIWTENGQSYIYKLPASCLPASDSFRSDV 300
QY 301 GKAVENLIIPVQHILDRKDKELLICPPVTRPPYGCREFPHKLLIQDSSGRINIWNISD 360
DB 301 GKAVENLIIPVQHILDRKDKELLICPPVTRPPYGCREFPHKLLIQDSSGRINIWNISD 360
QY 361 TADKQSGEGLMTTISIQEAFDKNPCPAGIIDQLSVTPNSNEPLKVTASVYIPAHGR 420
DB 361 TADKQSGEGLMTTISIQEAFDKNPCPAGIIDQLSVTPNSNEPLKVTASVYIPAHGR 420
QY 421 LVCGREDGSIIVIPATQTAIVQLQGEHMLRGWPPHRTLGRHNKVTCLLYPHQVSARY 480
DB 421 LVCGREDGSIIVIPATQTAIVQLQGEHMLRGWPPHRTLGRHNKVTCLLYPHQVSARY 480
QY 481 DQRYLLSGGVDFSVIWDIFSGEMKHI FCHVGGEITQLLVPPENCARSARVHCICVSADH 540
DB 481 DQRYLLSGGVDFSVIWDIFSGEMKHI FCHVGGEITQLLVPPENCARSARVHCICVSADH 540
QY 541 SVGLLSLRKCKITMLASRLHFPQVTKWRPSDDLVVVGCSGDSVYVQMDTGALDRCVNG 600
DB 541 SVGLLSLRKCKITMLASRLHFPQVTKWRPSDDLVVVGCTDGSVCVQMDTGALDRCVNG 600

QY 601 ITAVEILNACDEAVPAVDLSLHPAVNLQAMTRRSIALAKNMAHKLQTLATNLASEA 660
DB 601 ITAVEILNACDEAVPAVDLSLHPAVNLQAMTRRSIALAKNMAHKLQTLATNLASEA 660
QY 661 SDKGNLPKYSNLSMVQAIKTNLTDPDIHVLFPDVEALIIQLLTERASRNTALISPENL 720
DB 661 SDKGNLPKYSNLSMVQAIKTNLTDPDIHVLFPDVEALIIQLLTERASRNTALISPENL 720
QY 721 QKASGSDSGSFLTGKRAAVLFPQVKETIKENIKHLLDDEDEBEIMQRREESDPEY 780
DB 721 QKASGSDSGSFLTGKRAAVLFPQVKETIKENIKHLLDDEDEBEIMQRREESDPEY 780
QY 781 RSKSKPLFLELYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHCTVSPGLLSR 840
DB 781 RSKSKPLFLELYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHCTVSPGLLSR 840
QY 841 GGHMSLMLPGYNOPACKLSHGKTEVGRKLPASBGVCKGTGVSVRAVTTQHLLSIIISLANT 900
DB 841 GGHMSLMLPGYNOPACKLSHGKTEVGRKLPASBGVCKGTGVSVRAVTTQHLLSIIISLANT 900
QY 901 LMSMTNATFIGDHMKGPTRPPRPTDLSKARGSPPTSSNIVQGOIKOVAAPVVSARS 960
DB 901 LMSMTNATFIGDHMKGPTRPPRPTDLSKARGSPPTSSNIVQGOIKOVAAPVVSARS 960
QY 961 ADHSGSDPPSPALHTCFLVNEGWSQLAAHCVMLPDLGLGDKFRPPLLEMLARRWQDRC 1020
DB 961 ADHSGS-ASAPALHTCFLVNEGWSQLAAHCVMLPDLGLGDKFRPPLLEMLARRWQDRC 1020
QY 1021 LEVREAAQALLAELEIRIEQAGRKDAIDAWAPVLPQYIDHVI SPGVTSEAQTITAPDA 1080
DB 1021 LEVREAAQALLAELEIRIEQAGRKDAIDAWAPVLPQYIDHVI SPGVTSEAQTITAPDA 1080
QY 1079 SGPEAKVQSEEHDLVDDDDITTCGLSSVPQMKKISTSYEERRRKOATAVLLGVIGABFGAE 1140
DB 1079 SGPEAKVQSEEHDLVDDDDITTCGLSSVPQMKKISTSYEERRRKOATAVLLGVIGABFGAE 1140
QY 1141 IBPPKLLTRPRSSQIPEGFGLTSGSNYSLARHTCKALTFLLLQPPSPKLPHPHSTIRRT 1200
DB 1141 IBPPKLLTRPRSSQIPEGFGLTSGSNYSLARHTCKALTFLLLQPPSPKLPHPHSTIRRT 1200
QY 1199 ADLGLRGFTVWEPYMDVSAVLMGLLELCADAEKOLANITMGLPLSPAADSARSARHALS 1260
DB 1199 ADLGLRGFTVWEPYMDVSAVLMGLLELCADAEKOLANITMGLPLSPAADSARSARHALS 1260
QY 1261 LIATARPAPAFITIAKEVHRHTALAAANTOSQNMHTTTTLARAKEGLRVEILIEKOPTD 1320
DB 1261 LIATARPAPAFITIAKEVHRHTALAAANTOSQNMHTTTTLARAKEGLRVEILIEKOPTD 1320
QY 1321 VDILLVEVMDIIMYCLEGLVKKKGLQECFPAICRPFYVSYERNHRIAVGARHGSVALY 1380
DB 1321 VDILLVEVMDIIMYCLEGLVKKKGLQECFPAICRPFYVSYERNHRIAVGARHGSVALY 1380
QY 1381 DIRTGKCTQTHGKGPIITAVAPDGRYLATYNTDTSHTSFQWMTSLGSGTGLMSAPQ 1440
DB 1381 DIRTGKCTQTHGKGPIITAVAPDGRYLATYNTDTSHTSFQWMTSLGSGTGLMSAPQ 1440
QY 1441 LFCIKTYQVPPVQSPASPSHNAKLARLITWTSNRNVLMAHDKGKEHFRV 1490
DB 1439 LFCIKTYQVPPVQSPASPSHNAKLARLITWTSNRNVLMAHDKGKEHFRV 1488

RESULT 6

ABB66447

ID ABB66447 standard; protein; 1525 AA.

AC ABB66447;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 26133.

DE Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR N-PSDB; ABL10550.
XX DR MPI; 2001-656860/75.
XX DR N-PSDB; ABL10550.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX PT interactions.
XX PS Disclosure; SEQ ID NO 26133; 21pp + Sequence Listing; English.
XX PS The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signaling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB5773-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1525 AA;

Query Match 40.2%; Score 3128.5; DB 4; Length 1525;
Best Local Similarity 41.0%; Pred. No. 6e-278;
Matches 655; Conservative 303; Mismatches 458; Indels 183; Gaps 32;

QY 1 MAGNSIVLPVLVGRKAPTHCISAVLLTDDGATVTCGHDQICLWDLS-VELQINPRAL 59
DB 1 MVTNLVVPVLMGPTAPTHCISVFLSDDQFLVTGCVGQICLQVPEPTILKMPRL 60
QY 60 LFGHTASITCLSKACASSDKQYIVASBESGEMCLWDVSGRCIEFTKLACTHTGIIQYQF 119
DB 61 LVGHSAPVLCVRLPENNPLVSSSENGEMCTWDLTDGKMEAVKLPQVHTQIQSYH- 119
QY 120 SVGNQREGRLCHGHYPETILVVDATSLVLYSLVSKISPDWISSMSIIRSHRTQEDTVVA 179
DB 120 -TANSEDRVLCIGYAEIMVDPFSLVLYVLSKYPDPWISAIHVLPRMKRDKDVLA 178
QY 180 LSVTGLKWLIVTSETSDMDTEPIPEERSKPIYCONCOSISPCAFQSRLLVVCISKYWR 239
DB 179 ITTGTGVKWLITG--NENHAPPIYENESKEIRCLNATMNCQAQNRVLVLCYKIQW 236
QY 240 VFDAGDYSLCSPSENGQTWTGDFVSSDKVIITWENGOSYIYKLPASCLPASDSFRSD 299
DB 237 IYDAGDFTVLCVIAPARERWQGGDFITSDRVMLWTDGKGLYKLPANCI PDKNEFHS- 295
QY 300 VGKAVENLIPVQHILLDRDKELLICPPVTRFFYCGREYFPHKLLTQDSSGRINW--- 356
DB 296 --KSVVRDAPLYLVQLHAGDK-VLSCPPAMKLLQAGGQHN--LLRGDSEGYISVWNV 350
QY 357 -----NLSADADQGSSEGLAMTTSLSQAEAFKLANPCPAGIIDLQSVIPNSNEPLKVT 411
DB 351 EVELDNLISIIQAQMPRLKPHVCTSLVEANIMDPPVGIIDQLSRITES--PVKLTS 408
QY 412 SVYIPAGRLVCGEDGSGIVVPATQTAIVQLLQGEHMLRRGPPHRTLFGHRNKVTCLL 471
DB 409 SIYLPQOSRLVIGEDGSGIVVPATQTVMMQLLVGIKQNFSDWFSHQILYGHGRVNCLL 468

QY 472 YPHQVSARYDORYLISGGVDRSVIITWIDIFSGEMKHFVHGGEITQLLVPENCARVOH 531
DB 469 CPSMTHSRYEKSHLUSGGIDPAVCLWDLYSGSLHFRFCVHAGEITQLLVPENCSPRIUK 528
QY 532 CICSVASDHSVGLLSLREKCKCIMLASRHLFPQIKWRPSDDYLIVVGCSDGSGVYVQWMDT 591
DB 529 CICSVASDHSVTLVSLQERKCVTLASRHLFPVVTIKWRPLDDFLIVGSDGSGVYVQWMBT 588
QY 592 GALDRCVMGITAVEILNACDE-----AVPAADVLSHPAVNLQAMTRRSL 637
DB 589 GHLDRLVHGLMAEVLSCDEQAEDGGGGGSGNGASASEMGMANPAVHFFRGKSRNM 648
QY 638 AALKMAHHKLOTLATNLLASASDKGN--LPKYSHNSLMVQAIKTNLTDPDIDHVLFPD 694
DB 649 NAIR-----HATORGITQLOLQGHQGNFDFLMKHSNPLVIGIURTPKDAESHILFPD 704
QY 695 VEALIIQLITEASRPNTALISPENLOKASGSSDKGSFLTGKRAAVLFQVVK----- 747
DB 705 IEGILFELHSEBYAQMTPTATLESIGVHLQNPDKGSMHLDASKKIGDPFNKVKNAVDVE 764
QY 748 ETIKENIKHEHLLDDEEEDIMR-----QRRESDDPYRSKSKPKPLTLLYVNLTM 797
DB 765 KILDKDKHGLVQKFKETEIYVEKVOAKVESLOKAVEPHEEQDLKSKIASKMEVTHYM 824
QY 798 DTAKLFMSCLHAWGLNEVLDEVCLDRGLMGLKPHCTVVSFGLLSRGGHMSLMLPGYNQ--- 853
DB 825 EYVQALLSLLHSGWGLDPLHDKWCFETRLGULLRPIVISYGVLSKAGYMSLLLTWQNNYAI 884
QY 854 -PACKL--SHGTEYGRKLPASEGV-----GKGTYGVSRAVTTQHLLSIISLIANTLMSMT 905
DB 885 PFGIQLPSSSKKRLPEELQRLHLEHTAVPSTRLHWELSTLTNTNHLALVAMNTLLSMS 944
QY 906 NATFTGDHMKGPTRPPRSTPDLSKARGSPPTSSNIVGQIKQVAAAPVVSARSADHSG 965
DB 945 AASFL-----PDSEKHK-----KLQRLA-----QRTDSTLSN 971
QY 966 SPSPSAPALHTCFLVNEGWSOLAAMHCVMLPD---LLGLDKFRPPLLELARWODRCLE 1022
DB 972 EEERELMAHHSIQKHAWSLLATHCFLLPKIEALEPKPKRQVENMKVQWHHCIE 1031
QY 1023 VREAAQALLABLRREQAGRKEADAMAPYLPQYIDHVISPGVTSEAAQTITAP----- 1078
DB 1032 IREAAQILLGELTRMGKGRQLVESWQYLPVLYTH--TEPIVGAQOQALALISQASGG 1089
QY 1079 -----DASGPEAKV-----QEEHDLVDDDDITTCCLSSVPQM 1110
DB 1090 AGSGSGNGGVGVGSGGSGAGSGSGVPGGDAHQDEYEEEBEIE-----I 1139
QY 1111 KKISTSYBERRKOATAIVLLGVIGAFGAIEI--EPPK-----LLTRPRS 1152
DB 1140 RPSLSSELUKRIQTAVILLGVIGAFQODISQESPNHRGSI SMATGANLTSGVAGGER 1199
QY 1153 SSQIPEGFGLTSGGSNYSLARHTKALTFLLQPPSPKLPHPHTIRRTAIDLIGRFTVM 1212
DB 1200 KSSVVEGFGIAN-----NLARLTSNALAHLVAPPSPKLPQVTPLRRAAIDLIGRFTVM 1254
QY 1213 EPMVDSAVLMGELLELCADAQKOLANIWMGLPLSPAADSARSARHLSLIATARPAPFIT 1272
DB 1255 EYFLDYSKVLGLLEISCEG-KAVPNLNYKPLTPQADACRTARHARLRIATARPAPFIT 1313
QY 1273 TTAKEVHRHTALAAANTQSQNMHT--TTLARAKEGLRIVIEILIEKMTDWDVLLVEVMDI 1331
DB 1314 TWAREVARNYVWQNAQSIINTPLTQSVLHAKGEILQCVEMLDKQSEIAGLLVEVMDI 1373
QY 1332 IMYCLEGSLVKKGLQECPPAICRFYVMSYERNHRIAVGARHGSVALYDITFGKQCTH 1391
DB 1374 ALHCVDNGLKNKGLAELCPALCKFNQISHCATQTRIAVANGSNGMLAIYELRQNKQCMIP 1433
QY 1392 GHKGPITAVAPADGRIYATYNTDISHISFWQNTSLLGSIQMLNSAPQLRCIKTYQVPP 1451
DB 1434 AHTHPITSLAFSPDGKGLVSVSACENRLSFWOTSTGMFG-LGQ-----SQTRCTKGYSTAP 1488

QY 1452 VQASPGSHVAKLARIWTSNRVILMAHDGKEHFMV 1490
Db 1489 IPDVS--RLNPMRLAKLWNNRTVTLMLADGSETRENV 1525

RESULT 7

ADSS96554
ID ADS96554 standard; protein; 1525 AA.

XX AC ADS96554;

XX DT 02-DEC-2004 (first entry)

XX DE Drosophila melanogaster protein, SEQ ID 175.

XX KW Insecticide; Antiparasitic; Anthelmintic.

XX OS Drosophila melanogaster.

XX PN W02004039999-A2.

XX PD 13-MAY-2004.

XX PF 08-AUG-2003; 2003WO-US024982.

XX PR 30-OCT-2002; 2002US-0422377P.

XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX PI Stam L, Kandar KP, Spana E, Bachmann J;

XX DR WPI; 2004-376203/35.

XX DR N-PSDB; ADS96553.

XX PT Identifying a compound that inhibits the activity of a protein for
PT Drosophila viability for use e.g., as insecticidal agent by expressing in
PT a recombinant host a DNA molecule to produce a protein essential for
PT Drosophila viability.

XX PS Claim 1; SEQ ID NO 175; 57pp; English.

XX CC The present invention relates to a method for identifying a compound that
XX inhibits the activity of a protein essential for Drosophila viability.
XX The method comprises: (a) expressing in a recombinant host a DNA sequence
XX encoding a protein essential for Drosophila viability; (b) testing
XX compounds suspected of having the ability to inhibit the activity of the
XX protein expressed in (a); and identifying a compound tested in (b) that
XX inhibits the activity of the protein. The method is useful in identifying
XX a compound that inhibits the activity of a protein essential for
XX Drosophila viability for use as insecticidal, ectoparasiticide,
XX antiparasitic, anthelmintic or acaracidal agent. The present sequence
XX is one such protein essential for Drosophila viability. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1525 AA;

Query Match 40.2%; Score 3128.5; DB 8; Length 1525;

Best Local Similarity 41.0%; Pred. No. 6e-278;

Matches 655; Conservative 303; Mismatches 458; Indels 183; Gaps 32;

QY 1 MAGNSLVPLVLMGRKAPTHICISAVLLTDDGATVTVGCHDQGLCLWDLVS-VELQINPRAL 59

Db 1 MWTNLVVPVLMGPATPHICISVFLSDQFLTVTGCDGQGLCLWQVEPTLKMSPRCL 60

QY 60 LFQHTASITCLSKACASSDKQYIVSASEGEMCLWDVSDGRCIEPTKLACTHTGQFYOF 119

Db 61 LVGHSAPVLCVVRASLLPENNPLVSSSENGEMCTWDLTDGKWEAVKLPQVHTQIQSYH- 119

QY 120 SVGNQREGRLCHGHPEILVVDATSLVLYSLVSKISPDWISMSIIRSHRTQEDTVVA 179

Db 120 -TANSEVDVRLFCIGYAEIWMVDPFSLVLEVTVLSSKVPDWISAIHVLPRMRKDDVILA 178

QY 180 LSVTGILKWIWITSEISDMQDTEPIFEESKPIYQNCOSISFCAPTQBSLLVVCISKYMR 239
Db 179 ITTGTGVKWTLTG--NENKHAEPYENESKEIRCLNAITMNCACQNRQTLVLCVTKYQ 236
QY 240 VPDAGDYSLLCSPSENGQTTGGDFVSSDKVLIITENGQSYIYKLPASCLPASDSFRSD 239
Db 237 IYDAGDFTVLCSVIAPARERWQGGDFITSDRVMLTDEKGIDYKLPANCIPONKGFHS- 295
QY 300 VGKAVENLIPPVQHILLDRKDKELIICPPVTRFFYGCBEYFHKLLIQGSSGSRNIW--- 356
Db 296 --KSVVRDAPLYLVYVLOHAGDK-VLSCPPAMKLLQAGGQHN--LARGUSEGVISVWNP 350
QY 357 ----NISDTAKQSGSEGLAMTTSISLOEAFDKLNPAGIIDQISVIPNSNEPKVTA 411
Db 351 EYPLDNISILQAKQMPPLKPHVCTSLVEASIMDPPVPGILDQLSRITES--PVKLTLS 408
QY 412 SVYIPAHGRLVCGREDGSIIVPATQTAIVOLLQGEHMLRRGPPHRTLRGNKTKCLL 471
Db 409 SIYLPQOSRLVIGREDGSIIVPATQTVMMQLLVGIKQNFSDMPFSDHQLYGHGRVNCCL 468
QY 472 YPHOVSARYDQRYLISGGVDVFSVIWIDIFSEGEMKHIFCVHGGBITOLLVPPENC SARVOH 531
Db 469 CFSMTHSRYEKSHLLSGGIDFVCLWDLVSGSLLHFRFCVHAGEITOLLVPPESCSPRLK 528
QY 532 CICSVASDHSVGLLSLREKKICMLASRHLFPIQVWKRPESDDYLVVVGCSGDSVYVWQMDT 591
Db 529 CICSVASDHSVTLVLSLOERKCVTLASRHLFPVVTIKWRPLDDFLIVGCSGDSVYVWQMET 588
QY 592 GALDRCVMGITAVEILNACDE-----AVPAVDSLSHPAVNLKQAMTRSL 637
Db 589 GHLDRLVHGMABEVLSDACDEAEDGGGGGSGNGASASAMGMAPVHFFRGLKSRNM 648
QY 638 AALKWMAHKLQTLATNLASASDKGN--LPKYSHNSLMVQAIKTNLTDPDIHVLFPD- 694
Db 649 NAIR-----HATRGITQLQOLQHQNGNDFLWKHRSNPLVQLGRTNPKDAESHILFPD 704
QY 695 VEALIIQLLTEASRPNTALISPENLQKASGSDGSGSFLTGKRAAVLFOQVK- 747
Db 705 IEGLIFFELHSEVAQWTPATLESGLVHLQNPDKGSKMHLDAKKIGDFFNKVKNAVDVE 764
QY 748 EYIKENIKHLLDDEEDBEIMR-----ORRESDPEYRSKSKPLTLLENLTM 797
Db 765 KILQDKHGLVQKFKETIEVKVQAKVESLQKAVEPHEEQDULKSIAKMEYTHVM 824
QY 798 DTAKLFMSCLHAGLNEVLDEVLDRGLMKHCTVSGLISRGHMSLMPLQYNQ- 853
Db 825 EVAQLLSLHSLWGLDPLHDKMCETRLGLRPIVPSYGLSKAGTMSLLPTWQNNYAI 884
QY 854 -PACKL--SHGKTEVGRKLPASEGV-----GKGTGVSRVAVTTQHLLSIISLANTLMSMT 905
Db 885 PFGIQLPSSSKKRPPEELQRLHLEHTAVFTSRHLWELSTLTITNHLALVAMSNLTLMS 944
QY 906 NATFTGDHMKGPTRPPRSTPDLSKARSGPTSSNIVOGQIKQVAAPVVSARSADHSG 965
Db 945 AASFL-----PDSEKHK-----KLQRLA-----QRTDSTLSN 971
QY 966 SDPPSAPALHTCFVNEGWSQLAAMHCVMPLD---LLGLDKRPPPLEMLARWRDCL 1022
Db 972 EBERBELMAHHSIQIKHAWSLATHHCFLLPKIEALEPKPKRPQVEMVVRKHQHCIE 1031
QY 1023 VREAQAQLLAEALRIEIOAGRKEADAMAPYLPQYIDHVISPGVTSEAAQTITAP- 1078
Db 1032 IREAAQIILLGELTMGKGRQLVESWAQYLPYLYTH--TEPIVGAQQOQALALISQASGG 1089
QY 1079 -----DASGPEAKV-----QEEHDLVDDDDITTCGLSVPQM 1110
Db 1090 AGSGSGGNGGVGVSGGGGAGSGGPGSGVPGDAHQDEYEEBEEI-----I 1139
QY 1111 KKISYSYERRRQATAIVLLGVIGAEFGAEI--EPPK-----LLTRPRS 1152
Db 1140 RKPSSLSELKRKQTTAVILLGVIGAEFGQDISOESPNHRGSIEMATGANLTSGVAGERR 1199

QY 1153 SSOIPRSGFLTSGSNYSLARHTCKALTFTLLQPPSKLPHPHSITRRITDILIGRFTVM 1212
Db 1200 KSSVVEFGTAN-----NLRLTSMALHLLYAPPSPKLPQYTPLRRAIDLIGRFTVM 1254
QY 1213 EPTMYDSAVLMGLLELCAEAKOLANITMGLPLSPAADSARSARHALSLIATARPAPFIT 1272
Db 1255 EPLYDVSXVLLGLEISCEG-KAVPNLYKLPLTPQADACRTARHALRLIATARPAFIT 1313
QY 1273 TIAKEVHRHTALAAQTOSQONMIT-TTLARAKGEILRVIELIEKPTDVEDLLVERVMDI 1331
Db 1314 TMAREVARNTMQNAQSIINTPLTQSVLHKAKGEIILQCVELMDKQSEIAGLLVERVMDI 1373
QY 1332 IMYCLEGLVKKGLOECRPAICRFYMWVSVERNHRIAVCARHGSVALYDIRTKGCOTIH 1391
Db 1374 ALHCVDGNEKNGRLAELCPAICFKNQISHCAQTRRIAVGANGSNLAIYELRQNKQMP 1433
QY 1392 GHKGPITAVAPADGRYLATYSNTDGHISFWQNTSLLSIGMNSAPQLRCIKTYQVPP 1451
Db 1434 AHTPITSLAFSPDGKYLVSYSCAENRLSPWQSTGMFG-LGQ-----SQRCTKGYSTAP 1488
QY 1452 VQPASPGSHNALKARLIWTSNRNVILMAHDGKEHRPMV 1490
Db 1489 IPDVS--RLNPMRLAKLVWNNRTVTMLADGSETRFNV 1525

RESULT 8

ADAM5388
ID ADAM5388 standard; protein; 559 AA.

XX ADAM5388;

XX 20-NOV-2003 (first entry)

XX Human protein, SEQ ID 2956.

XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease.

XX Homo sapiens.

XX EP1293569-A2.

XX 19-MAR-2003.

XX 21-MAR-2002; 2002EP-00006586.

XX 14-SEP-2001; 2001JP-00328381.

XX 24-JAN-2002; 2002US-0350435P.

XX (HELI-) HELIX RES INST.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX WPI; 2003-395539/38.

DR N-PSDB; ADA53749.

XX New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.

XX Claim 14; SEQ ID NO 2956; 205pp; English.

XX The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.

XX Sequence 559 AA;

Query Match 36.1%; Score 2805.5; DB 6; Length 559;
Best Local Similarity 94.3%; Pred. No. 6.3e-249; Indels 33; Gaps 1;
Matches 549; Conservative 0; Mismatches 0;
QY 526 SARVQHICSVASDHSVGLLSLREKCCIMLASRHLFPPIQVIKWRPDDYLVVGCSDGSVY 585
Db 11 SARVQHICSVASDHSVGLLSLREKCCIMLASRHLFPPIQVIKWRPDDYLVVGCSDGSVY 70
QY 586 VQMDTGALDRCVMGITAVEILNACDEAVPAAVDSLSHPAVNLKQAMTFRSLAALKNMAH 645
Db 71 VQMDTGALDRCVMGITAVEILNACDEAVPAAVDSLSHPAVNLKQAMTFRSLAALKNMAH 130
QY 646 HKLQTLATNLASEASDGNLPHYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIQLLPE 705
Db 131 HKLQTLATNLASEASDGNLPHYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIQLLPE 190
QY 706 EASRPNTALISPENLQKASGSSDKGSGFLTGKRAAVLFOQVKETIKENIKEHLDDDEED 765
Db 191 EASRPNTALISPENLQKASGSSDKGSGFLTGKRAAVLFOQVKETIKENIKEHLDDDEED 250
QY 766 BEIMRQRRESDPEYRSSKSKPLTLLLYNLTDQTKLPMSCILHAWGLNEVLDEVCIDRLG 825
Db 251 BEIMRQRRESDPEYRSSKSKPLTLLLYNLTDQTKLPMSCILHAWGLNEVLDEVCIDRLG 310
QY 826 MLKPHCTVSGLLSRGGHMSLMLPGYNOPACKLSHGKTEVGRKLPASEGVGKGTGVRSRA 885
Db 311 MLKPHCTVSGLLSRGGHMSLMLPGYNOPACKLSHGKTEVGRKLPASEGVGKGTGVRSRA 370
QY 886 VTTOHLLSIISLANTIMSMNTATFIGDHMKGPTRPPRSTPDLKARGSPPTSSNIVQG 945
Db 371 VTTOHLLSIISLANTIMSMNTATFIGDHMKGPTRPPRSTPDLKARGSPPTSSNIVQG 430
QY 946 QIKQVAAFPVVSARSADHSGSDPPSPAPALHTCFLVNEGWSQLAAHMCVMLPDLGLDKFR 1005
Db 431 QIKQ-----GWSQLAAHMCVMLPDLGLDKFR 457
QY 1006 PPLLEMLARRWQDRCLEVRERAAQALLLAEELRIEQAQRKEAIDAWAPYLPQYIDHVISPG 1065
Db 458 PPLLEMLARRWQDRCLEVRERAAQALLLAEELRIEQAQRKEAIDAWAPYLPQYIDHVISPG 517
QY 1066 VTSEAAQTITTAPDASGPEAKVQEEHDLVDDDDITTGCLSSV 1107
Db 518 VTSEAAQTITTAPDASGPEAKVQEEHDLVDDDDITTGCLSSV 559

RESULT 9

ADM04809

ID ADM04809 standard; protein; 775 AA.

XX ADM04809;

XX 20-MAY-2004 (first entry)

XX Human protein of the invention SEQ ID NO:3494.

XX human; gene therapy; diagnostic marker; pharmaceutical.

XX Homo sapiens.

XX EP1347046-A1.

XX 24-SEP-2003.

XX 12-APR-2002; 2002EP-00008400.

XX 22-MAR-2002; 2002JP-00137785.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231368P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 27-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.

PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-476159/51.

N-PSDB; ADM19384.

Isolated nucleic acid molecule encoding a channel/transporter protein is used in preventing, treating or ameliorating a medical condition.

Claim 11; SEQ ID NO 670; 809pp; English.

The invention relates to an isolated nucleic acid molecule encoding a channel/transporter protein or sequences at least 95% identical to a these. The nucleic acids and proteins encoded by them are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathologic condition. The antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunoassays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia, angioneitis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. This sequence corresponds to a protein of the invention.

XX Sequence 209 AA;

Query Match 12.9%; Score 1003; DB 4; Length 209;

Best Local Similarity 98.5%; Pred. No. 2.8e-83;

Matches 191; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MAGNSLVLPFVLWGRKAPTHCISAVLLTDDGATVTVGCHDQICLMDLSVLEIQINPRALL 60
|||||

Db 11 MAGNSLVLPVWGRKAPTHCISAVLLTDDGATVTCCHDQICLMDLSVELQINPRALL 70
Qy 61 FGHATSTCLSKACASSDKQYIVSASSGEMCLWDVSDGRCEIETFKLACTHTTGIFQFYQFS 120
Db 71 FGHATSTCLSKACASSDKQYIVSASSGEMCLWDVSDGRCEIETFKLACTHTTGIFQFYQFS 130
Qy 121 VGNOREGRLLCHGHYPILVVDATSLVLYSLVSKISPDWISSMIIIRSHRTQEDTIVVAL 180
Db 131 VGNOREGRLLCHGHYPILVVDATSLVLYSLVSKISPDWISSMIIIRSHRTQEDTIVVAL 190
Qy 181 SVTGILKVMWIVTSE 194
Db 191 SVTGILKVMWIVTSE 204

RESULT 11
ADM19833
ID ADM19833 standard; protein; 210 AA.
XX
AC ADM19833;
XX
DT 20-MAY-2004 (first entry)
DE Protein encoded by novel human channel/transporter gene #151.
XX
KW immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
KW cytotatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW gene therapy; channel/transporter protein; rheumatoid arthritis;
KW neoplasm; cardiac arrest; cerebrovascular disorder; cerebral ischemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease;
KW ocular disorder; corneal infection; wound healing;
KW epithelial cell proliferation; skin aging; sunburn; transplantation;
KW chemotaxis; food additive.
XX
OS Homo sapiens.
XX
XX WO200154472-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001307.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214986P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 23-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228242P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 05-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241826P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.

PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0254097P.
PR 03-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476159/51.
DR N-PSDB; ADM19354.
XX

Isolated nucleic acid molecule encoding a channel/transporter protein is used in preventing, treating or ameliorating a medical condition.

Claim 11; SEQ ID NO 640; 809pp; English.

The invention relates to an isolated nucleic acid molecule encoding a channel/transporter protein or sequences at least 95% identical to a these. The nucleic acids and proteins encoded by them are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. The antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia, angiodenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. This sequence corresponds to a protein of the invention.

Sequence 210 AA;

Query Match 11.7%; Score 909; DB 4; Length 210;
Best Local Similarity 100.0%; Pred. No. 1.4e-74;

Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1302 AKGEILRVIEILIEKPTDVVDLLVEVMDIIMYCLEGLVKKKGLQECPPAICRFYVVS 1361
DB 1 AKGEILRVIEILIEKPTDVVDLLVEVMDIIMYCLEGLVKKKGLQECPPAICRFYVVS 60
QY 1362 YERNHRIAVGARHGSVALYDITGKQTIHGKGPITAVAFAPDGRYLATYNTDSHISF 1421
DB 61 YERNHRIAVGARHGSVALYDITGKQTIHGKGPITAVAFAPDGRYLATYNTDSHISF 120
QY 1422 WQNTSLLGSGIGWLSAPQLCIKTYQVPPVPASPFGSHNALKLARLIWTSNR 1474
DB 121 WQNTSLLGSGIGWLSAPQLCIKTYQVPPVPASPFGSHNALKLARLIWTSNR 173

RESULT 12

ADM20106

ID ADM20106 standard; protein; 81 AA.

XX ADM20106;

XX 20-MAY-2004 (first entry)

DE Protein encoded by novel human channel/transporter gene #181 clone 2.

KW immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
KW cytoskeletal; cardiac; vasotropic; cerebroprotective; neurotropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW gene therapy; channel/transporter protein; rheumatoid arthritis;
KW neoplasm; cardiac arrest; cerebrovascular disorder; cerebral ischemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease;
KW ocular disorder; corneal infection; wound healing;
KW epithelial cell proliferation; skin aging; sunburn; transplantation;
KW chemotaxis; food additive.

XX Homo sapiens.

XX WO200154472-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001307.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 11-JUL-2000; 2000US-0217496P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0225213P.

XX 14-AUG-2000; 2000US-0225214P.

XX 14-AUG-2000; 2000US-0225266P.

XX 14-AUG-2000; 2000US-0225267P.

XX 14-AUG-2000; 2000US-0225268P.

XX 14-AUG-2000; 2000US-0225270P.

XX 14-AUG-2000; 2000US-0225447P.

XX 14-AUG-2000; 2000US-0225757P.

XX 14-AUG-2000; 2000US-0225758P.

XX 14-AUG-2000; 2000US-0225759P.

PR	18-AUG-2000;	2000US-02262799P
PR	22-AUG-2000;	2000US-02266611P
PR	22-AUG-2000;	2000US-02266688P
PR	23-AUG-2000;	2000US-02271182P
PR	23-AUG-2000;	2000US-02271700P
PR	30-AUG-2000;	2000US-02289247P
PR	01-SEP-2000;	2000US-02292877P
PR	01-SEP-2000;	2000US-02293437P
PR	01-SEP-2000;	2000US-02293438P
PR	01-SEP-2000;	2000US-02311242P
PR	01-SEP-2000;	2000US-02293444P
PR	05-SEP-2000;	2000US-02293454P
PR	05-SEP-2000;	2000US-02295097P
PR	08-SEP-2000;	2000US-02295137P
PR	08-SEP-2000;	2000US-02304377P
PR	08-SEP-2000;	2000US-02311137P
PR	08-SEP-2000;	2000US-02311414P
PR	08-SEP-2000;	2000US-02320808P
PR	08-SEP-2000;	2000US-02320811P
PR	12-SEP-2000;	2000US-02311968P
PR	14-SEP-2000;	2000US-02323977P
PR	14-SEP-2000;	2000US-02323988P
PR	14-SEP-2000;	2000US-02323999P
PR	14-SEP-2000;	2000US-02324001P
PR	14-SEP-2000;	2000US-02330637P
PR	14-SEP-2000;	2000US-02330647P
PR	21-SEP-2000;	2000US-02342237P
PR	21-SEP-2000;	2000US-02342747P
PR	25-SEP-2000;	2000US-02349977P
PR	25-SEP-2000;	2000US-02349987P
PR	26-SEP-2000;	2000US-02354847P
PR	27-SEP-2000;	2000US-02358347P
PR	27-SEP-2000;	2000US-02358367P
PR	29-SEP-2000;	2000US-02363277P
PR	29-SEP-2000;	2000US-02363767P
PR	29-SEP-2000;	2000US-02365687P
PR	29-SEP-2000;	2000US-02365697P
PR	29-SEP-2000;	2000US-02363707P
PR	02-OCT-2000;	2000US-02370377P
PR	02-OCT-2000;	2000US-02370387P
PR	02-OCT-2000;	2000US-02370397P
PR	02-OCT-2000;	2000US-02370407P
PR	13-OCT-2000;	2000US-02399357P
PR	13-OCT-2000;	2000US-02399377P
PR	20-OCT-2000;	2000US-02405960P
PR	20-OCT-2000;	2000US-02411217P
PR	20-OCT-2000;	2000US-02417857P
PR	20-OCT-2000;	2000US-02417867P
PR	20-OCT-2000;	2000US-02418087P
PR	20-OCT-2000;	2000US-02418267P
PR	01-NOV-2000;	2000US-02445177P
PR	08-NOV-2000;	2000US-02464747P
PR	08-NOV-2000;	2000US-02464757P
PR	08-NOV-2000;	2000US-02464767P
PR	08-NOV-2000;	2000US-02464777P
PR	08-NOV-2000;	2000US-02464787P
PR	08-NOV-2000;	2000US-02452337P
PR	08-NOV-2000;	2000US-02452527P
PR	08-NOV-2000;	2000US-02452527P
PR	08-NOV-2000;	2000US-02452567P
PR	08-NOV-2000;	2000US-02452577P
PR	08-NOV-2000;	2000US-02452827P
PR	08-NOV-2000;	2000US-02455327P
PR	08-NOV-2000;	2000US-02466097P
PR	08-NOV-2000;	2000US-02465107P
PR	08-NOV-2000;	2000US-02466117P

PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249276P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250319P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
XX	(HUMA-)	HUMAN GENOME SCI INC.
PA	Rosen CA,	Barash SC, Ruben
PI	WPI; 2001-476159/51.	N-PSDB; ADM19627.
DR	Isolated nucleic acid molecu	used in preventing, treating
XX	Claim 11; SEQ ID NO 913;	809pp
PS	The invention relates to an i	channel/transporter protein o
CC	these. The nucleic acids and	prevent, treat or ameliorate
CC	rabbits, goats, horses, cats,	in diagnosing a pathological
CC	pathological condition. The a	in alleviating symptoms assoc
CC	immunoassays e.g. radiomun	(ELISA). Disorders which are
CC	diseases e.g. rheumatoid arth	neoplasms of the breast or li
CC	arrest, cerebrovascular disor	nervous system disorders e.g.
CC	bacteria, viruses and fungi a	The polypeptides can also b
CC	cell proliferation, to preven	organs before transplantation
CC	tissues, to regenerate tissue	also be used as a food additi
CC	storage capabilities. This se	invention.
XX	Sequence 81 AA;	
SQ		

Query Match 4.3%; Score 332; DB 4; Length 81;
Best Local Similarity 89.9%; Pred. No. 4.5e-22;

Matches 62; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MAGNSLVPLVLMGRKAPTHCISAVLLTDGATIVTGCHDQICLWDLSEVQLINPRALL 60
 |||||
 Db 11 MAGNSLVPLVLMGRKAPTHCISAVLLTDGATIVTGCHDQICLWDLSEVCKLLILQTL 70
 |||||

QY 61 FGHASITC 69
 |||||
 Db 71 FGHASITC 79
 |||||

RESULT 13
 ADS29860
 ID ADS29860 standard; protein; 579 AA.
 XX AC ADS29860;
 XX XX
 DT 02-DEC-2004 (first entry)
 XX DE Bacterial polypeptide #18893.
 XX
 KW Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polypeptide.
 XX OS Bacteria.
 XX US2003233675-A1.
 XX PD 18-DEC-2003.
 XX PF 20-FEB-2003; 2003US-00369493.
 XX PR 21-FEB-2002; 2002US-0360039P.
 XX PA (CAOY/) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.
 XX PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 XX WPI; 2004-061375/06.
 XX
 PT New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.
 XX
 PS Claim 1; SEQ ID NO 18893; 122pp; English.
 XX
 CC The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by

CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polypeptide used in the
 CC scope of the invention. Note: The sequence data for this patent did not
 CC form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 579 AA;

Query Match 2.9%; Score 228.5; DB 8; Length 579;
 Best Local Similarity 20.6%; Pred. No. 6.1e-11;
 Matches 120; Conservative 80; Mismatches 203; Indels 179; Gaps 24;

QY 19 THCISAVLLTDGATIVTGCHDQICLWDLSEVQLINPRALLFGHTASITCLSKACASDQ 78
 |||||
 Db 109 TSCVRSVFPSSDGAWLASGSDQTVRLWDIS---SGNCLYTLQGH---SCVRSVFPSPD 162
 |||||

QY 79 KOYIVSASESGEMCLMDVSDGRCIBFTKLACTHTGTQFYQFVSGNQREGRLCHGHYPI 138
 |||||
 Db 163 GAWLASGGDQIVRLWDISSGNCL-YTLQGYT-SWVRFVFS---PNGVTLANGSSDQI 216
 |||||

QY 139 L-VVDATSLVLYSLVSKISPDWISSMSIIRSHRTOEDTVVALSVTG--ILKWIVTSBI 195
 |||||
 Db 217 VRLWDISSKKCLYTLQG--HTNWNVAVP-----SPDGATLASGSDQTVRLWDISS-- 266
 |||||

QY 196 SDMQDTEPIFEBESKPIYQCNQCSISFCAFTQRSLLVWCCKYV---RVFDAGDYSLLCSG 252
 |||||
 Db 267 -----SKCLY-----ILQHTSWNSVFNPD-DGSTLASG 295
 |||||

QY 253 PSENGQWTGDFVSSDKVI-IWTENGQSYIYKLPASCLPASDSFRSDYKAVENLIPPV 311
 |||||
 Db 296 -----SSDQTVRLWEINS-----SKCLCTFQGHTSWNSVFN----- 328
 |||||

QY 312 QHILLRDKKELLICPPVTRPFYGCREFYFKLLIOGSSGRLNINWISDTADKQSEGL 371
 |||||
 Db 329 -----PD-----GSMASGSDKTVRLWDISSKCLHTFQGH 361
 |||||

QY 372 AMTTSISLQAFDKLNPAGIIDQLSVIPNSNEPLKVTVSVIYPAHGLRVLGREGDGV 431
 |||||
 Db 362 NWVNSVA-----FNP-----DGSM 376
 |||||

QY 432 IV-PATQTAIVQLQGEHMLRGPPTLRGHRNKVTCLLYPHQVSARYDQRYLISGGV 490
 |||||
 Db 377 ASGSGDQTVRLWEISSKCL-----HTFQGHTSWNS-----SVTPSPDGTMLASGSD 423
 |||||

QY 491 DFSVIWIDIFSGEMKHIFCVHGBETQLLVPPENCARSVQHICISVASHVGLLSLREK 550
 |||||
 Db 424 DQTVRLWISSGECCLYTLFGHTNWWGVSIVFSPDGA-----TLASGSDQTVRLWISSG 477
 |||||

QY 551 KCIMLASRHLFPQIVTKWRPSPDDYLWVGCSGVSVMQMDTG 592
 |||||
 Db 478 KCLYTLQGHNNWVGVSIVFSPDGTLLASGSDQTVRLWISSG 519
 |||||

RESULT 14
 ADS41794
 ID ADS41794 standard; protein; 608 AA.
 XX AC ADS41794;
 XX XX
 DT 02-DEC-2004 (first entry)
 XX DE Bacterial polypeptide #20224.
 XX
 KW Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polypeptide.
 XX OS Bacteria.

XX US2003233675-A1.
XX 18-DEC-2003.
XX 20-FEB-2003; 2003US-00369493.
XX 21-FEB-2002; 2002US-0360039P.
XX (CAOY/) CAO Y.
XX (HINK/) HINKLE G J.
XX (SLAT/) SLATER S C.
XX (CHEN/) CHEN X.
XX (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
XX PT for expression of a polynucleotide encoding a polypeptide from a
XX PT microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 20224; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
XX CC promoter functional in a plant cell, where the promoter is positioned to
XX CC provide for expression of a polynucleotide encoding a polypeptide from a
XX CC microbial source. The invention also relates to a transformed plant
XX CC comprising the recombinant DNA construct and a method of producing a
XX CC transformed plant having an improved property. The plant is a crop plant
XX CC such as maize or soybean. The method of producing a transformed plant
XX CC having an improved property comprises transforming a plant with the
XX CC recombinant DNA construct and growing the transformed plant, where the
XX CC polynucleotide or polypeptide is useful for improving plant properties.
XX CC The recombinant DNA construct is useful for producing plants with
XX CC improved plant properties, e.g. improved cold, heat or drought tolerance,
XX CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX CC increased resistance to plant disease, better growth rate by modification
XX CC of the cell cycle pathway with plant growth regulators, increased rate of
XX CC homologous recombination, modified seed oil or protein yield and/or
XX CC content, improved yield by modification of carbohydrate, nitrogen or
XX CC phosphorus use and/or uptake, by modification of photosynthesis or by
XX CC providing improved plant growth and development under at least one stress
XX CC condition, improved lignin production or improved galactomannan
XX CC production. This sequence represents a bacterial polypeptide used in the
XX CC scope of the invention. Note: The sequence data for this patent did not
XX CC form part of the printed specification but was obtained in electronic
XX CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX SQ Sequence 608 AA;

Query Match 2.9%; Score 228.5; DB 8; Length 608;
Best Local Similarity 18.4%; Pred. No. 6.7e-11;
Matches 112; Conservative 89; Mismatches 176; Indels 231; Gaps 25;

QY 23 SAVLLTDDGATVITGCHDQICLWDLV-ELQINPRALLFGHTASITCLSCACASDKQY 81
DB 27 AGTAFSPDGTLLATGABGELRLWEVATGKLVN-----FAGHLGWWSLA---FSPDQQL 79
QY 82 IVSASSEGCMWDSGRCIEFTKLACTHTGTFQYQFSVGNQREGRL-----130
DB 80 LASCSDKTIKRLWVNTGKCLR-----TLSGHTSIWSVAFSADQOMLASGDEPTIRL 133
QY 131 -----CH-----GHYPEILVDATSLVLYSLVSKISPDWISSMIIIRSHRTQETVVA 179
DB 134 WNVNTGDKHKIPSGHTRILSL-----SPSSDGQTLAS 166
QY 180 LSVTGLKLVWTVTSEISDQDTEPIEERSKPIYQCNCOSISCAFTQSRLLVWCYSKYR 239
DB 167 GSADFIRLWKISGE-----CDRILEGHSRIW-----SISF-----198
QY 240 VFDAGDYSLLCGSPSENGQTWTGGDFVSSDKVI-IWTENGQSYIYKLPASCLPASPSFRS 298

DB 199 -----SPDQQLVSG-----SADFTIRLWEVS-----TGNCF-----225
QY 299 DVGKAVENLIPVQHILLDKD--KELLICPPVTRFFYCGREYFHKLLIQDSSGRLNIW 356
DB 226 -----NILQHSRDRSLAFSPA-----QMLYSASDDKTVRIW 259
QY 357 NISDTADKQGSBEGLAMTTSISLQEAFLKLNFCPPAGIIDOLSVIP-NSNEPLKVTASVYI 415
DB 260 EAS-----TGEC-----LNILPGHTNSIFSVAFNV--284
QY 416 PAHGRVLCGREGSIVIVPATQTAIVQLQGEHMLRGWPPH-----RTLGRHNKVTCL 470
DB 285 -----DGRTIASGSTDTV-----KLWDVNTGRCFKTLKGSNSV--319
QY 471 LYPHOVSARYDORYLISGGVDFSPVIVWIDIFSGEMKHIFCVHGGEITQLLVPPENCARSVQ 530
DB 320 ---FSVAFNLDQQLASGSTDTQTVRLMDVNTGCLKKFAGHSGWTVSVAFHPDG-----370
QY 531 HCICSVASDPHVSGLLSLREKKCIMLASRHLFPPIQVIKWRPSDDYLVVGCSDGSVYVWQMD 590
DB 371 DELLASSADRTIRLWSVSTGQCLQLKDHVNVQSVAFSPDRQILASGSDDDQIRLWSVS 430
QY 591 TQALDRCV 598
DB 431 TG---KCL 435

RESULT 15
ADS41812
ID ADS41812 standard; protein; 1005 AA.
XX
XX ADS41812;
XX
XX 02-DEC-2004 (first entry)
XX Bacterial polypeptide #20242.
XX
XX Recombinant DNA construct; transformed plant; improved plant property;
XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
XX pathogen tolerance; pest tolerance; plant disease resistance;
XX cell cycle pathway modification; plant growth regulator;
XX homologous recombination; seed oil yield; protein yield; carbohydrate;
XX nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
XX bacterial polypeptide.
XX Bacteria.
XX
XX US2003233675-A1.
XX PD 18-DEC-2003.
XX
XX 20-FEB-2003; 2003US-00369493.
XX
XX 21-FEB-2002; 2002US-0360039P.
XX
XX (CAOY/) CAO Y.
XX (HINK/) HINKLE G J.
XX (SLAT/) SLATER S C.
XX (CHEN/) CHEN X.
XX (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
XX PT for expression of a polynucleotide encoding a polypeptide from a
XX PT microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 20242; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 17:12:11 ; Search time 215 Seconds
(without alignments)
3548.828 Million cell updates/sec

Title: US-10-645-335-2

Perfect score: 7773

Sequence: 1 MAGNSLVLPVLWGRKAPTH.....TSNRNVILMAHDGKHFV 1490

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7773	100.0	1490	2 Q86UX5	Q86ux5 homo sapien
2	7573.5	97.4	1457	2 Q96P87	Q96p87 homo sapien
3	7408.5	95.3	1489	2 Q92019	Q92019 mus musculus
4	7373	94.9	1488	2 Q9ERH3	Q9erh3 rattus norv
5	7254.5	93.3	1457	2 Q920J0	Q920J0 mus musculus
6	7247.5	93.2	1457	2 Q920I8	Q920I8 rattus norv
7	6023	77.5	1160	1 WDR7 HUMAN	Q9y4e6 homo sapien
8	5784.5	74.4	1166	2 Q80TY3	Q80ty3 mus musculus
9	4331	55.7	956	2 Q6GPQ7	Q6gpq7 xenopus lae
10	3170	40.8	614	2 Q86VP2	Q86vp2 homo sapien
11	3128.5	40.2	614	2 Q9W543	Q9w543 drosophila
12	3107.5	40.0	1473	2 Q7Q3S0	Q7q3s0 anopheles g
13	2944	37.9	1471	2 Q46092	Q46092 drosophila
14	2667	34.3	531	2 Q8C711	Q8c711 mus musculus
15	2006.5	25.8	1335	2 Q9BL05	Q9bl05 caenorhabdi
16	1610	20.7	1102	2 Q7Z3I3	Q7z3i3 homo sapien
17	1442	18.6	775	2 Q8N8X2	Q8n8x2 homo sapien
18	839	10.8	343	2 Q96OM9	Q96om9 drosophila
19	798	10.3	1342	2 Q86L37	Q86l37 dictyosteli
20	518.5	6.7	1446	2 Q9SUT4	Q9sut4 arabidopsis
21	331	4.3	117	2 Q8BPW7	Q8bpw7 mus musculus
22	234	3.0	1526	1 YV46 ANASP	Q8yrl1 anabaena sp
23	230.5	3.0	1227	2 Q8Z0R1	Q8zor1 anabaena sp
24	229	2.9	1258	1 YS00 ANASP	Q8ytc2 anabaena sp
25	223	2.9	1184	2 Q7ND85	Q7nd85 gloeobacter
26	221.5	2.8	1081	2 Q7NCT8	Q7nct8 gloeobacter
27	217.5	2.8	1592	2 Q62471	Q62471 caenorhabdi
28	206.5	2.7	1923	2 Q6AWF2	Q6awf2 drosophila
29	205.5	2.6	1193	2 Q7ND05	Q7nd05 gloeobacter
30	204	2.6	1680	2 Q9UAQ2	Q9uaq2 caenorhabdi
31	203.5	2.6	1601	2 Q8WSM1	Q8wsm1 caenorhabdi

ALIGNMENTS

RESULT 1

Q86UX5

ID Q86UX5 PRELIMINARY; PRT; 1490 AA.

AC Q86UX5;

DT 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Rabconnectin-3 beta.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22671656; PubMed=12786944;

RA Kawabe H., Sakisaka T., Yasumi M., Shingai T., Izumi G., Nagano F.,

RA Deguchi-Tawarada M., Takeuchi M., Nakanishi H., Takai Y.,;

RT "A novel rabconnectin-3-binding protein that directly binds a GDP/GTP

RT exchange protein for Rab3A small G protein implicated in Ca(2+)-

RT dependent exocytosis of neurotransmitter.";

RL Genes Cells 8:537-546(2003).

CC -1- SIMILARITY: Contains 6 WD repeats.

DR EMBL; AY093325; AM33134.1; -

DR InterPro; IPR001680; WD40.

DR InterPro; IPR011046; WD40_like.

DR Pfam; PF00400; WD40; 6.

DR PRINTS; PR00320; GPROTEINBRPT.

DR SMART; SM00320; WD40; 7.

DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_2.

DR PROSITE; PS50082; WD_REPEATS_2; 4.

DR PROSITE; PS50294; WD_REPEATS_REGION; 3.

KW Repeat; WD repeat.

SQ SEQUENCE 1490 AA; 163809 MW; 626E876245696A1P CRC64;

Query Match 100.0%; Score 7773; DB 2; Length 1490;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1490; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGNSLVLPVLWGRKAPTHCSAVLLTDDGATVVGCHDGOICLWDLVELQINPRALL 60

Db 1 MAGNSLVLPVLWGRKAPTHCSAVLLTDDGATVVGCHDGOICLWDLVELQINPRALL 60

QY 61 FGHASITCLSKACASSDKQYIVSASEGEMCLWDVSDGRCIEFTKLACTHTGTGQYQFS 120

Db 61 FGHASITCLSKACASSDKQYIVSASEGEMCLWDVSDGRCIEFTKLACTHTGTGQYQFS 120

QY 121 VGNQREGRLCHGHYPILVVDATSLVLYSVLSKISPDWISSMSIIRSHRQEDTVVAL 180

Db 121 VGNQREGRLCHGHYPILVVDATSLVLYSVLSKISPDWISSMSIIRSHRQEDTVVAL 180

QY 181 SVTGILKWIVTSEISDMQDTEPIFEESKPIYCONCQSIISFCAPTORSLVVCYSKYWRV 240

Db 181 SVTGILKWIVTSEISDMQDTEPIFEESKPIYCONCQSIISFCAPTORSLVVCYSKYWRV 240

```
QY 241 FDAGDYSLLCSPGSENGQWTGDFVSSDKVIITWENGQSYIYKLPASCLPASDSFRSDV 300
DB 241 FDAGDYSLLCSPGSENGQWTGDFVSSDKVIITWENGQSYIYKLPASCLPASDSFRSDV 300
QY 301 GRAVENLIPPVQHILLDRKDKELLICPPVTRFFYGCREFYFHKLLIQGDSGRLNIWNISD 360
DB 301 GRAVENLIPPVQHILLDRKDKELLICPPVTRFFYGCREFYFHKLLIQGDSGRLNIWNISD 360
QY 361 TADKQSEGLAMTTSISLQEAFLKLPASCLPASDSFRSDV 420
DB 361 TADKQSEGLAMTTSISLQEAFLKLPASCLPASDSFRSDV 420
QY 421 LVCGRDGSIVIVPATQTAIQLLQGEHMLRGWPPHRTLRGRNKNVTCLLYPHOVASARY 480
DB 421 LVCGRDGSIVIVPATQTAIQLLQGEHMLRGWPPHRTLRGRNKNVTCLLYPHOVASARY 480
QY 481 DQRYLISGGVDRSVIIVDIFSGEMKHFVHGGEITQLLVPPENCARSARVQHCICSVASDH 540
DB 481 DQRYLISGGVDRSVIIVDIFSGEMKHFVHGGEITQLLVPPENCARSARVQHCICSVASDH 540
QY 541 SVGLLSLRKKCIMLASRLHFPQVIKWRPSDDYLVGCSGDSGVYVQMDTGALDRCVNG 600
DB 541 SVGLLSLRKKCIMLASRLHFPQVIKWRPSDDYLVGCSGDSGVYVQMDTGALDRCVNG 600
QY 601 ITAVEILNACDEAVPAVDSLHPA VNLQAMTRRSALALKNMAHKLQTLATNLLASEA 660
DB 601 ITAVEILNACDEAVPAVDSLHPA VNLQAMTRRSALALKNMAHKLQTLATNLLASEA 660
QY 661 SDKGNLPKYSHNSLWQAIKTNLTDPDIHVLFPDVEALIIQLLTBEASRNTALISPENL 720
DB 661 SDKGNLPKYSHNSLWQAIKTNLTDPDIHVLFPDVEALIIQLLTBEASRNTALISPENL 720
QY 721 QKASGSDSGSFLTKRAAVLFPQVKETIKENIKHLLDDEEDEBEIMRQRESDPY 780
DB 721 QKASGSDSGSFLTKRAAVLFPQVKETIKENIKHLLDDEEDEBEIMRQRESDPY 780
QY 781 RSKSKPLTLLEVLNMTDAKLFMSCLHAWGLNEVLDEVLDRGLMKPHCTVSGFLLSR 840
DB 781 RSKSKPLTLLEVLNMTDAKLFMSCLHAWGLNEVLDEVLDRGLMKPHCTVSGFLLSR 840
QY 841 GGHMSLMLPGYNQPACKLSHGKTEVGRKLPASEGVGKGTGVGSRAVTTQHLLSIISLANT 900
DB 841 GGHMSLMLPGYNQPACKLSHGKTEVGRKLPASEGVGKGTGVGSRAVTTQHLLSIISLANT 900
QY 901 LMSMTNATIGDHMKKGPTRPPSTPDLISKARGSPPTSSNIVQGOIKOVAAPVVSARD 960
DB 901 LMSMTNATIGDHMKKGPTRPPSTPDLISKARGSPPTSSNIVQGOIKOVAAPVVSARD 960
QY 961 ADHSGSDPPSAPALHTCFVLNMGWSQLAAHCVMLPDLLGLDKFRPPLLEMLARWQDR 1020
DB 961 ADHSGSDPPSAPALHTCFVLNMGWSQLAAHCVMLPDLLGLDKFRPPLLEMLARWQDR 1020
QY 1021 LEVREAAQALLAELRRIEQAQRKEAIDAWAPYLPQYIDHVTSPGVTSEAAQTITAPDA 1080
DB 1021 LEVREAAQALLAELRRIEQAQRKEAIDAWAPYLPQYIDHVTSPGVTSEAAQTITAPDA 1080
QY 1081 SGPEAKVQEEHDLVDDDDITTCCLSSVPQMKKISTSYEERRKOATAI VLLGVIGAFEGAE 1140
DB 1081 SGPEAKVQEEHDLVDDDDITTCCLSSVPQMKKISTSYEERRKOATAI VLLGVIGAFEGAE 1140
QY 1141 IEPKLLTTPRSSQIPEGFGLTSGGSNTSLARHTCKALTFLLOPPSPKLPHPHSTIR 1200
DB 1141 IEPKLLTTPRSSQIPEGFGLTSGGSNTSLARHTCKALTFLLOPPSPKLPHPHSTIR 1200
QY 1201 AIDLIGRGFTVMEPYMDVSAVLMGLLELCADEKQALANITMGLPLSPAADSARSARHALS 1260
DB 1201 AIDLIGRGFTVMEPYMDVSAVLMGLLELCADEKQALANITMGLPLSPAADSARSARHALS 1260
QY 1261 LIATARPPAFITIAKEVHRHTALAAANTOSQONMHTTTLARAKGEILRVIIEKMPD 1320
DB 1261 LIATARPPAFITIAKEVHRHTALAAANTOSQONMHTTTLARAKGEILRVIIEKMPD 1320
QY 1321 VVDLLVEVMDIIMYCLEGSLVKKKLQECFPAICRFYVMVSYERNHRIA VAGRHGSVALY 1380
```

```
DB 1321 VVDLLVEVMDIIMYCLEGSLVKKKLQECFPAICRFYVMVSYERNHRIA VAGRHGSVALY 1380
QY 1381 DIRTOKCQTIHGHKGPITAVAPADGRYLATYNTDSDSHISFWQNTSLGSGMLNSAQ 1440
DB 1381 DIRTOKCQTIHGHKGPITAVAPADGRYLATYNTDSDSHISFWQNTSLGSGMLNSAQ 1440
QY 1441 LRCIKTYQVPPVOPASPGSHNALKARLIWTSNRNVILMAHDGKEHFMV 1490
DB 1441 LRCIKTYQVPPVOPASPGSHNALKARLIWTSNRNVILMAHDGKEHFMV 1490

RESULT 2
Q96PS7 PRELIMINARY; PRT; 1457 AA.
AC Q96PS7;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE TGF-beta resistance-associated protein TRAG (Fragment).
GN Name=TRAG;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RC Tissue=Liver;
RA Sanders S., Thorgeirsson S.S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: Contains 6 WD repeats.
EMBL: AF188125; AAL03983.1; -.
DR InterPro: IPR001680; WD40.
DR InterPro: IPR011046; WD40-like.
DR Pfam: PF00400; WD40; 6.
DR PRINTS: PR00320; GPROTEINRPT.
DR SMART: SM00320; WD40; 7.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.
DR PROSITE: PS00082; WD_REPEATS_2; 4.
DR PROSITE: PS00294; WD_REPEATS_REGION; 3.
KW Repeat; WD repeat.
FT NON_TER 1457
SQ SEQUENCE 1457 AA; 160494 MW; 202B6A1C5EB0E2F7 CRC64;

Query Match 97.4%; Score 7573.5; DB 2; Length 1457;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 1456; Conservative 1; Mismatches 0; Indels 33; Gaps 1;

QY 1 MAGNSLVLPVLWGRKAPTHCISAVLLTDDGATIVTGCHDGOICLWDLSELQINPRALL 60
DB 1 MAGNSLVLPVLWGRKAPTHCISAVLLTDDGATIVTGCHDGOICLWDLSELQINPRALL 60
QY 61 FGHASITCLSKACASSDKQYIVSASEGEMCLWVSDGRCIEFTKLACTHTGQIYQPS 120
DB 61 FGHASITCLSKACASSDKQYIVSASEGEMCLWVSDGRCIEFTKLACTHTGQIYQPS 120
QY 121 VGNQREGRLCHGHYPEILLVVDATSLVLYSVLSKISPDWISSMSTIRSHRTOEDTVVAL 180
DB 121 VGNQREGRLCHGHYPEILLVVDATSLVLYSVLSKISPDWISSMSTIRSHRTOEDTVVAL 180
QY 181 SVTGILKVMIVTSEISDMQDTEPIFEESKPIYCONCQISIFCAFTQSRLLVCSKYWRV 240
DB 181 SVTGILKVMIVTSEISDMQDTEPIFEESKPIYCONCQISIFCAFTQSRLLVCSKYWRV 240
QY 241 FDAGDYSLLCSPGSENGQWTGDFVSSDKVIITWENGQSYIYKLPASCLPASDSFRSDV 300
DB 241 FDAGDYSLLCSPGSENGQWTGDFVSSDKVIITWENGQSYIYKLPASCLPASDSFRSDV 300
QY 301 GRAVENLIPPVQHILLDRKDKELLICPPVTRFFYGCREFYFHKLLIQGDSGRLNIWNISD 360
DB 301 GRAVENLIPPVQHILLDRKDKELLICPPVTRFFYGCREFYFHKLLIQGDSGRLNIWNISD 360
QY 361 TADKQSEGLAMTTSISLQEAFLKLPASCLPASDSFRSDV 420
```


Db 361 TADKQSGSEGLAMTTSISLQAFDKNPCPAGIIDQSVIPNSNEPLKVTASVYI PAHGR 420
QY 421 LVCGRDGSIVIPATQTAIVQLLQGEHMLRRGWPPHRTLGRHNKVTCLLYPHQVSARY 480
Db 421 LVCGRDGSIVIPATQTAIVQLLQGEHMLRRGWPPHRTLGRHNKVTCLLYPHQVSARY 480
QY 481 DQRYLISGGVDFSVIWIIDFSGEMKHIFCVHGGEITQLLPPENC SARVOHCICSVASDH 540
Db 481 DQRYLISGGVDFSVIWIIDFSGEMKHIFCVHGGEITQLLPPENC SARVOHCICSVASDH 540
QY 541 SVGLLSREKCKTCLASRLHFPPIQVWKRPDDVLLVVGCSGSGVYVWMDTGALDRCVMG 600
Db 541 SVGLLSREKCKTCLASRLHFPPIQVWKRPDDVLLVVGCSGSGVYVWMDTGALDRCVMG 600
QY 601 ITAVEILNACDEAVPAVDLSHPAVNLKQAMTERRSLAALKNMAHKKLQTLATNLLASEA 660
Db 601 ITAVEILNACDEAVPAVDLSHPAVNLKQAMTERRSLAALKNMAHKKLQTLATNLLASEA 660
QY 661 SDKGNLPKYSHNSLMQAIKTNLTDPDIHVLFFDVEALIIQLLTHEASRPNTALISPENL 720
Db 661 SDKGNLPKYSHNSLMQAIKTNLTDPDIHVLFFDVEALIIQLLTHEASRPNTALISPENL 720
QY 721 QKASGSDKGSFLLTGKRAAVLFOQVKETIKENIKHLLDDEDEBEIMRQREESDPEY 780
Db 721 QKASGSDKGSFLLTGKRAAVLFOQVKETIKENIKHLLDDEDEBEIMRQREESDPEY 780
QY 781 RSSKSPKPLTLLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLELGLMKPHCTVSGFLLSR 840
Db 781 RSSKSPKPLTLLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLELGLMKPHCTVSGFLLSR 840
QY 841 GGHMSLMLPCYNQPACKLSHGKTEVGRKLPASBGVKGKTVGSRVTTQHLISII SLANT 900
Db 841 GGHMSLMLPCYNQPACKLSHGKTEVGRKLPASBGVKGKTVGSRVTTQHLISII SLANT 900
QY 901 LMSMTNATPFGDHMKGPTRPPRSTPDLKARGSPPTSSNIYQGIKQVAAVPVARS 960
Db 901 LMSMTNATPFGDHMKGPTRPPRSTPDLKARGSPPTSSNIYQGIKQVAAVPVARS 960
QY 961 ADHSGSDPPSPALHTCFLNCGWSOLAAMHCVMPLDGLGLDFRPPLEMLARRWQDRC 1020
Db 961 ADHSGSDPPSPALHTCFLNCGWSOLAAMHCVMPLDGLGLDFRPPLEMLARRWQDRC 987
QY 1021 LEVREAAQALLAEALRRIEQAGKEAIDAWAPYLPQYIDHVISPGVTSEAAQTITAPDA 1080
Db 988 LEVREAAQALLAEALRRIEQAGKEAIDAWAPYLPQYIDHVISPGVTSEAAQTITAPDA 1047
QY 1081 SGPEAKVQEEHDLVDDDDITGCLSSVPQMKKISTSYEERKQATAILVLLGVIGAEFGAE 1140
Db 1048 SGPEAKVQEEHDLVDDDDITGCLSSVPQMKKISTSYEERKQATAILVLLGVIGAEFGAE 1107
QY 1141 IEBPKLLTRPRSSQIPEGGLTSGGSNYSLARHTCKALTFLLLQPPSPKLPPhSTIRRT 1200
Db 1108 IEBPKLLTRPRSSQIPEGGLTSGGSNYSLARHTCKALTFLLLQPPSPKLPPhSTIRRT 1167
QY 1201 AIDLIGRGFTWEPYMDVSAVLMGLLELCADEKQALANITMGLPLSPAADSARSARHALS 1260
Db 1168 AIDLIGRGFTWEPYMDVSAVLMGLLELCADEKQALANITMGLPLSPAADSARSARHALS 1227
QY 1261 LIATARPPATITIAKEVHRHTALANTOSQONMHTTLARAKEILRVIEILIEKMPDT 1320
Db 1228 LIATARPPATITIAKEVHRHTALANTOSQONMHTTLARAKEILRVIEILIEKMPDT 1287
QY 1321 VVDLLVEVMDIIMYCLEGSLVKKGLQECPPALCRFYMVSYYERNHRIAVGARHGVSVALY 1380
Db 1288 VVDLLVEVMDIIMYCLEGSLVKKGLQECPPALCRFYMVSYYERNHRIAVGARHGVSVALY 1347
QY 1381 DIRTGKQTHGHKGPITAVAPADGRYLATYNTDISHISFWQMTSLLSGIGMANSAPQ 1440
Db 1348 DIRTGKQTHGHKGPITAVAPADGRYLATYNTDISHISFWQMTSLLSGIGMANSAPQ 1407
QY 1441 LRCIKTYQVPPVQSPGSHNALKLARLIWTSNRNVLMAHDKGHRFMV 1490

Db 1408 LRCIKTYQVPPVQSPGSHNALKLARLIWTSNRNVLMAHDKGHRFMV 1457
RESULT 3
Q92019 PRELIMINARY; PRT; 1489 AA.
AC Q92019; SEQUENCE FROM N.A.
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
TGFBeta resistance-associated protein TRAG (Fragment).
DE Name=Wdr7; Synonyms=Trag;
GN Mus musculus (Mouse);
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP STRAIN=1298VJ; TISSUE=Brain;
RC Sanders S.; Thorgeirsson S.S.;
RA Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 6 WD repeats.
DR EMBL; AF188124; AAL03982.1; -;
DR MGD; MGI:1860197; Wdr7.
DR Pfam; PF00400; WD40; 6.
DR PRINTS; PR00320; GPROTEINRPT.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_3.
DR PROSITE; PS00682; WD_REPEATS_2; 4.
DR PROSITE; PS00294; WD_REPEATS_REGION; 3.
KW Repeat; WD repeat.
FT NON TER 1489 1489
SQ SEQUENCE 1489 AA; 163387 MW; 6B50C7958F0D7940 CRC64;
Query Match 95.3%; Score 7408.5; DB 2; Length 1489;
Best Local Similarity 94.9%; Pred. No. 0;
Matches 1414; Conservative 36; Mismatches 39; Indels 1; Gaps 1;
QY 1 MAGNSLVLPIVLWGRKAPTHCISAVLLTDDGATVVGCHDQICLWLSVLELQINPRALL 60
Db 1 MAGNSLVLPIVLWGRKAPTHCISAVLLTDDGATVVGCHDQICLWLSVLELQINPRALL 60
QY 61 FGHASITCLSKACASDQYIVSASEGEMCLWVSDGRCIEFTKLACTHTGIGYQFS 120
Db 61 FGHASITCLSKACASDQYIVSASEGEMCLWVSDGRCIEFTKLACTHTGIGYQFS 120
QY 121 VGNORGRLLCHGHYPEILLVVDATSLVLYSLVSKISPDWISSMSIIRSHRTQEDTVVAL 180
Db 121 VGNORGRLLCHGHYPEILLVVDATSLVLYSLVSKISPDWISSMSIIRSHRTQEDTVVAL 180
QY 181 SVTGILKWIWITSEISDMQDTEPIFEESKPIYQNCQSIQSFCAFTQSLVVCSSKYRW 240
Db 181 SVTGILKWIWITSEISDMQDTEPIFEESKPIYQNCQSIQSFCAFTQSLVVCSSKYRW 240
QY 241 FDAGDYSLLCSPGSENGQWTGDFVSSDKVIITWENGQSYIYKLPAASDPSRSDV 300
Db 241 FDAGDYSLLCSPGSENGQWTGDFVSSDKVIITWENGQSYIYKLPAASDPSRSDV 300
QY 301 GRAVENLIIPVQHILLDRKOKELLICPPVTRFPYGCREFYHKLIIQDSSGRINIWIAD 360
Db 301 GRAVENLIIPVQHILLDRKOKELLICPPVTRFPYGCREFYHKLIIQDSSGRINIWIAD 360
QY 361 TADKQSGSEGLAMTTSISLQAFDKNPCPAGIIDQSVIPNSNEPLKVTASVYI PAHGR 420
Db 361 TADKQSGSEGLAMTTSISLQAFDKNPCPAGIIDQSVIPNSNEPLKVTASVYI PAHGR 420
QY 421 LVCGRDGSIVIPATQTAIVQLLQGEHMLRRGWPPHRTLGRHNKVTCLLYPHQVSARY 480
Db 421 LVCGRDGSIVIPATQTAIVQLLQGEHMLRRGWPPHRTLGRHNKVTCLLYPHQVSARY 480
QY 481 DQRYLISGGVDFSVIWIIDFSGEMKHIFCVHGGEITQLLPPENC SARVOHCICSVASDH 540
Db 481 DQRYLISGGVDFSVIWIIDFSGEMKHIFCVHGGEITQLLPPENC SARVOHCICSVASDH 540

Qy	541	SVGLLSREKKCI MLASRHLFP IQVI KWPRSDDYLVVGCSDG SVYVQMDTGALDRCVMG	600
Db	541	SVGLLSREKKCI MLASRHLFP IQVI KWPRSDDYLVVGCSDG SVYVQMDTGALDRCVMG	600
Qy	601	ITAVEIILNACDEAVPAADVSLSHPAVNLQAMTRRS LAALKNAHKKLQTLATNLILASEA	660
Db	601	ITAVEIILNACDEAVPAADVSLSHPAVNLQAMTRRS LAALKNAHKKLQTLATNLILASEA	660
Qy	661	SDKGNLPKYSHNSLMWQAIKTNLTDPIHLFPDVEALIIQLLTHEASRENTALISPENL	720
Db	661	SDKGNLPKYSHNSLMWQAIKTNLTDPIHLFPDVEALIIQLLTHEASRENTALISPENL	720
Qy	721	QKASGSDKGSGFLTGKRAAVLFOQVKETIKENIKEHLDDDEDEDEIMQREESDPEY	780
Db	721	QKASGSDKGSGFLTGKRAAVLFOQVKETIKENIKEHLDDDEDEDEARQRESDPEY	780
Qy	781	RSSKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPCHTVSPGLLSR	840
Db	781	RASKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPCHTVSPGLLSR	840
Qy	841	GCHMSLMLPGYNQPACKLSHGKTEVGRKLPASRGVKGKTYGSRVTTQHLLSIISLANT	900
Db	841	GCHMSLMLPGYNQAAKGKLLHAKAEVGRKLPAAEGVGKGYTVSGRAVTTQHLLSIISLANT	900
Qy	901	LMSMTNATFGDHHMKGPTPPRPSTPDLKARGSPPTSSNI VOGQIKOVAA PVWSARD	960
Db	901	LMSMTNATFGDHHMKGPTPPRPSTPDLKARDSPASSNI VOGQIKQAAAPVWSARD	960
Qy	961	ADHSGSDPPAPALHTCFLVNEGWSQALAAHCVMLPDLGLGLDKFRPPLLEMLARRWQDRC	1020
Db	961	ADHSGSDSAS- PALPTCFLVNEGWSQALAAHCVMLPDLGLGLERFRPPLLEMLARRWQDRC	1019
Qy	1021	LEVREAAQALLAELRRIEIQAGRKEADAWAPYLPOVIDHVI SPGVTSAAQITITPADDA	1080
Db	1020	LEVREAAQALLAELRRIEIQAGRKETITDTPAPYLPQYMDHVI SPGVTAEAQMTWAAAPDA	1079
Qy	1081	SGPPAKVQEEHDLVDDDIITGCLSSVPQMKKISTSYEERRKQATAVILGVIGABEGAE	1140
Db	1080	SGPPAKVQEEHDLVDDDIITAGCLSSVPQMKKISTSYEERRKQATAVILGVIGABEGAE	1139
Qy	1141	IEPPKLLTRPRSSQIIEPGFGLTSGGSNYSLARHTCKALTFLLLQPPSPKLPPHSTIRRT	1200
Db	1140	IEPPKLLTRPRSSQIIEPGFGLTSGGSNYSLARHTCKALTYLLLLQPPSPKLPPHSTIRRT	1199
Qy	1201	AIDLIGRGFTWPEYMDVSAVLMGLLELCADA EKQLANITWGLPLSPAADSARSARHALS	1260
Db	1200	ATDILIGRGFTWBPYMDVSAVLMGLLELCADA EKQLANITWGLPLSPAADSARSARHALS	1259
Qy	1261	LIATARPPAFITTTAKEVHRHTALAA NTQSQNMHTTTILARAKEILIRVTEIILIEKMPDT	1320
Db	1260	LIATARPPAFITTTAKEVHRHTALAA NTQSQIHTTTILARAKEILIRVTEIILIEKMPDT	1319
Qy	1321	VVDLLVEVMDIIMVCLSGSLVKKGLOECPAICRFYVMSYERNHRIAVGARHGSVALY	1380
Db	1320	VVDLLVEVMDIIMVCLSGSLVKKGLOECPAICRFYVMSYERSHRIAVGARHGSVALY	1379
Qy	1381	DIRTKGQCTIGHKGPITAVAFADPGRYLATYSNTDISHISFWQNTSLLGSI GMLNSAPO	1440
Db	1380	DIRTKGQCTIGHKGPITAVSPADGRYLATYSNTDISHISFWQNTSLLGSI GMLNSAPO	1439
Qy	1441	LRCIKTYQVPPVOPASPGSHNALKLARLIWTSNRRNVILMAHDGKEHRFMV	1490
Db	1440	LRCIKTYQVPPVOPASPGSHNALKLARLIWTSNRRNVILMAHDGKEHRFMV	1489

DE	TGF-beta resistance-associated protein.
GN	Name=TRAG;
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX	NCBI_TaxID=10116;
RN	[1]
RF	SEQUENCE FROM N.A.
RP	TISSUE=Brain;
RC	Sanders S., Thorgeirsson S.S.;
RA	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RL	-/- SIMILARITY: Contains 6 WD repeats.
CC	EMBL; AF305813; AA031140.1; -.
DR	InterPro; IPR001680; WD40.
DR	InterPro; IPR011046; WD40_like.
DR	Pfam; PF004400; WD40; 6.
DR	PRINTS; PR00320; GPROTEINERPT.
DR	SMART; SM00320; WD40; 6.
DR	PROSITE; PS00678; WD REPEATS 1; UNKNOWN_3.
DR	PROSITE; PS50082; WD REPEATS 2; 4.
DR	PROSITE; PS50294; WD_REPEATS_REGION; 3.
KW	Repeat; WD repeat.
SK	SEQUENCE 1488 AA; 163189 MW; A675B5A2F3C09803 CRC64;
QY	Query Match 94.9%; Score 7373; DB 2; Length 1488;
DB	Best Local Similarity 94.5%; Pred. No. 0;
DB	Matches 1408; Conservative 39; Mismatches 41; Indels 2; Gaps 2;
QY	1 MAGNSLVLPVLWGRKAPHCISAVLLTDDGATIVTGCDDGQICLDWLSVELQINPRALL 60
DB	1 MAGNSLVLPVLWGRKAPHCISILLTDDGGITVTCGDDGQICLDWLSVELQINPRALL 60
QY	61 FGHASTITCLSKACASDDKQYIVSASGESMCLWDYSDGRCIEFTKLACTHTGICQFYQFS 120
DB	61 FGHAAITCLSKACASDDKQYIVSASANGEMCLWDVNDGRCIEFTKLACTHTGICQFYQFS 120
QY	121 VGNQREGRLCHGHYPILLVVDATSLVLYSVLSKISPDWISSMSIIRSHRTQEDTVVAL 180
DB	121 VGNQREGRLCHGHYPILLVVDATSLVLYSVLSKISPDWISSMSIIRSHRTQEDTVVAL 180
QY	181 SVTGILKWIIVTSEISDMQDTEIPFEESKPIYQNCQSIISFCAFTQSRLLVVCVKYRW 240
DB	181 SVTGILKWIIVTSEISGLQDTEIPFEESKPIYQNCQSLSFCAFTQSRLLVVCVKYRW 240
QY	241 FDAGDYSLLCSGSENGQWTGDDFYSSDKVLIWTENGOSYIYKLPASCLPASDSFRSDV 300
DB	241 FDAGDYSLLCSGSENGQWTGDDFYSDKVIWTENGOSYIYKLPASCLPASDSFRSDV 300
QY	301 GKAVENLIPVQHILDRKDKELLI CPPVTRPFYGCREYFHKLLIQDSSGRLLNNISD 360
DB	301 GKAVENLIPVQHSLLDQKRELVI CPPVTRPFYGCCKEYVHLKLLIQDSSGRLLNNIAD 360
QY	361 TADKQSEGLAMWTTISIQEAFDKLNPAGIIDLQSVIPGNEPLKVTASVYIIPAHGR 420
DB	361 IADKQEANGLKTTTICSLQDAFDKLPKCPAGIIDLQSVIPNSNEPLKVTASVYIIPAHGR 420
QY	421 LVCGREGDSIVIPATQTAIVQLLQGEHMLRRCGWPPHRTLGRHNKVTCLLYPHQVSARY 480
DB	421 LVCGREGDSIIIVIPATQTAIVQLLQGEHMLRRCGWPPHRTLGRHNKVTCLLYPHQVSARY 480
QY	481 DQRYLISGGVDFSVIITWDFSGEMKHFVCHGGEITQLLVPPENC SARVOHCICSVASDH 540
DB	481 DQRYLISGGVDFSVIITWDFSGEMKHFVCHGGEITQLLVPPENC SARVOHCVCVASDH 540
QY	541 SVGLLSLREKCKTMLASRLHFPQVTKWRPSDDYLVGCGSDGSVYVQMDTGALDRVMG 600
DB	541 SVGLLSLREKCKTMLASRLHFPQVTKWRPSDDYLVGCGSDGSVYVQMDTGALDRVMG 600
QY	601 ITAVEIINLACDEAVPAVDSLSHPAVNLQAMTFRSLAALKNAWHKLTQLATNLLASEA 660
DB	601 ITAVEIINLACDEAVPAVDSLSHPAVNLQAMTFRSLAALKNAWHKLTQLATNLLASEA 660
QY	661 SDKGNLPKPYSHNLSIMVQAIKTNLTDPDIHVLFPDFEALIIQLLITEASRNTALISPENL 720

```
Db 661 SDKGNLPKYSNLSMVAQIKTNLTDPDIHVLFFDVEALIIQLLTERASRNTALISPENL 720
Qy 721 QKAGSGSDKGGSLTGTGRAAVLFOQVKETIKENIKHLLDDEDEDEIMRQREESDPEY 780
Db 721 QKAGSGSDKGGSLTGTGRAAVLFOQVKETIKENIKHLLDDEDEDEIMRQREESDPEY 780
Qy 781 RSKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRLGMLKPHCTVSGFLLSR 840
Db 781 RASKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRLGMLKPHCTVSGFLLSR 840
Qy 841 GGHMSLWLPYQNPQACKLSHGKTEVGRKLPASGEVGGKTYGVSRAVTTQHLISILANT 900
Db 841 GGHMSLWLPYQNPQACKLSHGKTEVGRKLPASGEVGGKTYGVSRAVTTQHLISILANT 900
Qy 901 LMSMTNATFIDGDMKGPTRPPRPTDLSKARGSPPTSSNIVQGOIKQVAAAPVWSARSD 960
Db 901 LMSMTNATFIDGDMKGPTRPPRPTDLSKARGSPPTSSNIVQGOIKQVAAAPVWSARSD 960
Qy 961 ADHSGSDPPSAPALHTCFLVNEGWSQLAAMCHVMLPDLGLDKFRPPLLEMLARRWQDRC 1020
Db 960 ADHSGS-ASASPALHTCFLVNEGWSQLAAMCHVMLPDLGLDKFRPPLLEMLARRWQDRC 1018
Qy 1021 LEVREAAQALLAELRIEAGKEAIDAWPYLPQYIDHVISPGVTSEAAQTITAPDA 1080
Db 1019 LEVREAAQALLAELRIEAGKEAIDAWPYLPQYIDHVISPGVTSEAAQTITAPDA 1078
Qy 1081 SGPEAKVQEEHDLVDDITTCGLSSVPQMKKISTSYEERRKQATVILGVTCABFGAE 1140
Db 1079 SGPEAKVQEEHDLVDDITTCGLSSVPQMKKISTSYEERRKQATVILGVTCABFGAE 1138
Qy 1141 IEPPKLLTRPRSSQIPEGFGLTSGGSNYSLARHTCKALFFLLQPPSPKLPHPSTIRRT 1200
Db 1139 IEPPKLLTRPRSSQIPEGFGLTSGGSNYSLARHTCKALFFLLQPPSPKLPHPSTIRRT 1198
Qy 1201 AIDLIGRFTVWEPYMDVSAVLMGLLELCADEKOLANITMGLPLSPAADSARSARHALS 1260
Db 1199 AIDLIGRFTVWEPYMDVSAVLMGLLELCADEKOLANITMGLPLSPAADSARSARHALS 1258
Qy 1261 LIATARPPAFTTIAKEVHRHTALAAANTQSQNMHTTTLARAKEILRVIELIEKMPDT 1320
Db 1259 LIATARPPAFTTIAKEVHRHTALAAANTQSQNMHTTTLARAKEILRVIELIEKMPDT 1318
Qy 1321 VDVLLEVMVDIIMYCLEGLSVKKGLEQCPAICRFYMWVSYRHRHIAVGARHGSVALY 1380
Db 1319 VDVLLEVMVDIIMYCLEGLSVKKGLEQCPAICRFYMWVSYRHRHIAVGARHGSVALY 1378
Qy 1381 DIRTGKQTHGKGPITAVAPADGRYLATYNTDHSISFWQMTSLIGSICMLNSAPQ 1440
Db 1379 DIRTGKQTHGKGPITAVAPADGRYLATYNTDHSISFWQMTSLIGSICMLNSAPQ 1438
Qy 1441 LRCIKTYQVPPVQSPGSHNALKLARLIWTSNRNVLMAHDCKEHFW 1490
Db 1439 LRCIKTYQVPPVQSPGSHNALKLARLIWTSNRNVLMAHDCKEHFW 1488

RESULT 5
Q920J0
ID Q920J0 PRELIMINARY; PRT; 1457 AA.
AC Q920J0;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE TGF-beta resistance-associated protein TRAG (Fragment).
GN Name=Wdr7; Synonyms=Trag;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SVJ; TISSUE=Liver;
RA Sanders S., Thorpeirsson S.S.;
```

```
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -f- Similarity: Contains 6 WD repeats.
DR EMBL; AF188123; AAL03981.1; -.
DR MGD; MGI:1860197; Wdr7.
DR Pfam; PF00400; WD40; 6.
DR PRINTS; PR00320; GPROTEINERPT.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_3.
DR PROSITE; PS0082; WD_REPEATS_2; 4.
DR PROSITE; PS0294; WD_REPEATS_REGION; 3.
KW Repeat; WD repeat.
FT NON_TER 1457 1457
SQ SEQUENCE 1457 AA; 160262 MW; EF33D0E44869E60A CRC64;

Query Match 93.3%; Score 7254.5; DB 2; Length 1457;
Best Local Similarity 93.0%; Pred. No. 0;
Matches 1386; Conservative 36; Mismatches 35; Indels 33; Gaps 1;

Qy 1 MAGNSLVLPVILWGRKAPTHCISAVLLTDDGATIVTGDGQICLDWLSVLEQINPRALL 60
Db 1 MAGNSLVLPVILWGRKAPTHCISAVLLTDDGATIVTGDGQICLDWLSVLEQINPRALL 60
Qy 61 FGHASITCLSKACASSDKOYIVSASESGEMCLWDVSDGRCIEFTKLACTHTGTGQYQFS 120
Db 61 FGHASITCLSKACASSDKOYIVSASESGEMCLWDVSDGRCIEFTKLACTHTGTGQYQFS 120
Qy 121 VGNQREGLLCHGHYPEILVVDATSLVLSVSKISPDWISSMSIIRSHRTQEDTVVAL 180
Db 121 VGNQREGLLCHGHYPEILVVDATSLVLSVSKISPDWISSMSIIRSHRTQEDTVVAL 180
Qy 181 SVTGILKWIIVTSBISDMQDTBPFIPEESKPIYQNCQISFCAPTORSLLVVCSKYWRV 240
Db 181 SVTGILKWIIVTSBISDMQDTBPFIPEESKPIYQNCQISFCAPTORSLLVVCSKYWRV 240
Qy 241 FQAGDYSLCSPSENGQOTWGTGDFVSDKVIITWENGOSYIYKLPASCLPASDSFRSDV 300
Db 241 FQAGDYSLCSPSENGQOTWGTGDFVSDKVIITWENGOSYIYKLPASCLPASDSFRSDV 300
Qy 301 GRAVENLIIPVQHILLDRDKKELLICPPVTRFPYGCYREYFHKLLIOGSSGRNLINWISD 360
Db 301 GRAVENLIIPVQHILLDRDKKELLICPPVTRFPYGCYREYFHKLLIOGSSGRNLINWISD 360
Qy 361 TADKQSEGLAWTTSISLOEAFDKLNPAGIIDQLSVIPSNEPLKVTASVYIIPAHGR 420
Db 361 TADKQSEGLAWTTSISLOEAFDKLNPAGIIDQLSVIPSNEPLKVTASVYIIPAHGR 420
Qy 421 LVCGREDGSIIVPATQTAIVOLLQGEHMLRGWPPHRTLGRHNKVTCLLYPHQVSARY 480
Db 421 LVCGREDGSIIVPATQTAIVOLLQGEHMLRGWPPHRTLGRHNKVTCLLYPHQVSARY 480
Qy 481 DORYLISGVDFSVIITWIDIFSGEMKHI FCHVGGEITQLLVPPENCARSARVQHICSVASDH 540
Db 481 DORYLISGVDFSVIITWIDIFSGEMKHI FCHVGGEITQLLVPPENCARSARVQHICSVASDH 540
Qy 541 SVGLLSLRKKKICMLASRHLFPQIVIKWRPSDDYLVVGCSDGSVYVQMDTGALDRCVMG 600
Db 541 SVGLLSLRKKKICMLASRHLFPQIVIKWRPSDDYLVVGCSDGSVYVQMDTGALDRCVMG 600
Qy 601 ITAVETLNACDEAVPAVDSLSHPAVNLKQATRSALAKMVAHKLQTLATNLASEA 660
Db 601 ITAVETLNACDEAVPAVDSLSHPAVNLKQATRSALAKMVAHKLQTLATNLASEA 660
Qy 661 SDKGNLPKYSNLSMVAQIKTNLTDPDIHVLFFDVEALIIQLLTERASRNTALISPENL 720
Db 661 SDKGNLPKYSNLSMVAQIKTNLTDPDIHVLFFDVEALIIQLLTERASRNTALISPENL 720
Qy 721 QKAGSGSDKGGSLTGTGRAAVLFOQVKETIKENIKHLLDDEDEDEIMRQREESDPEY 780
Db 721 QKAGSGSDKGGSLTGTGRAAVLFOQVKETIKENIKHLLDDEDEDEIMRQREESDPEY 780
Qy 781 RSKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRLGMLKPHCTVSGFLLSR 840
Db 781 RASKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRLGMLKPHCTVSGFLLSR 840
```

QY	841	GGHMSLMLPGYNOPACKLSHGKTEVGRKLPASBGVKGTYGVSRVTTQHLHSIISLANT	900
Db	841	GGHMSLMLPGYNQAAGKLHAKAEVGRKLPAAEGVGKGYTVSRVTTQHLHSIISLANT	900
QY	901	LMSMTNATFIGDHMKKGTPRPPRSTPDLISKARGSPPTSSNIVQGOIKQVAAPVVSARS	960
Db	901	LMSMTNATFIGDHMKKGTPRPPRSTPDLISKARDSPPASSNIVQGOIKQ	949
QY	961	ADHSGSDPPSAPALHTCFLVNEGWSQLAAMHCVMLPDLGLGDKFRPPLLEMLARRWQDRC	1020
Db	950	-----GWSQLAAMHCVMLPDLGLGLERFRPPLLEMLARRWQDRC	987
QY	1021	LEVREAAQALLAELARRIEQAGKEAIDAWAPVLPQYIDHVISPGVTSAAQTITTPADA	1080
Db	988	LEVREAAQALLAELARRIEQAGKETIDWAPVLPQYMDHVISPGVTSAAQTITTPADA	1047
QY	1081	SGPEAKVQBEEDHLDVDDITTCGLSSVPQMKKISTSYEERRKQATAIVLLGVIGAEFGAE	1140
Db	1048	SGPEAKVQBEEDHLDVDDITAGCLSSVPQMKKISTSYEERRKQATAIVLLGVIGAEFGAE	1107
QY	1141	IEPPKLLTPRSSSQIPGEGFLTSGGSNYSLARHTCKALTFLLOPPSPKLPHPSTIRRT	1200
Db	1108	IEPPKLLTPRSSSQIPGEGFLTSGGSNYSLARHTCKALTFLLOPPSPKLPHPSTIRRT	1167
QY	1201	AIDLIGRGFTWEPYMDVSAVLMGLLELCADAEKOLANTMGLPLSPAADSARSARHALS	1260
Db	1168	ATDLIGRGFTWEPYMDVSAVLMGLLELCADAEKOLANTMGLPLSPAADSARSARHALS	1227
QY	1261	LIATARPAPAFITIAKEVHRHTALAANTOSQNMHTTTLARAKGEILRVEILIEKMPDT	1320
Db	1228	LIATARPAPAFITIAKEVHRHTALAANTOSQSIHTTTLARAKGEILRVEILIEKMPDT	1287
QY	1321	VDLLVEVMDIIMYCLEGSLVKKKLQECFPAICRFYMWYSYERNHRIAVGARHGSVALY	1380
Db	1288	VDLLVEVMDIIMYCLEGSLVKKKLQECFPAICRFYMWYSYERSHRIAVGARHGSVALY	1347
QY	1381	DIRTGKQCTHGHKGPITAVAPADGRYLATYSNTDISHISFWQNTSLGSLGMLNSAPQ	1440
Db	1348	DIRTGKQCTHGHKGPITAVAPADGRYLATYSNTDISHISFWQNTSLGSLGMLNSAPQ	1407
QY	1441	LRCIKTYQVPPVQVAPSGSHNALKLARLIWTSNRNVILMAHDGKEHRFMV	1490
Db	1408	LRCIKTYQVPPVQVAPSGSHNALKLARLIWTSNRNVILMAHDGKEHRFMV	1457
RESULT 6			
ID	Q92018	PRELIMINARY; PRT; 1457 AA.	
AC	Q92018;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DE	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
DE	TCF-beta resistance-associated protein TRAG (Fragment).		
GN	Name=Trag;		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=101116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Liver;		
RA	Sanders S., Thorgeirsson S.S.;		
RL	Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.		
CC	-I- SIMILARITY: Contains 6 WD repeats.		
DR	EMBL; AF192379; AAL03984.1; -.		
DR	Pfam; PF00400; WD40; 6.		
DR	PRINTS; PR00320; GPROTEINERPT.		
DR	SMART; SM00320; WD40; 6.		
DR	PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_3.		
DR	PROSITE; PS00682; WD_REPEATS_2; 4.		
DR	PROSITE; PS0294; WD_REPEATS_REGION; 3.		
KW	Repeat; WD repeat.		
FT	NON TER	1457	1457
SQ	SEQUENCE	1457 AA;	160192 MW; 5E4AEAL3BCFF33B2 CRC64;
Query Match			
Best Local Similarity 93.2%; Score 7247.5; DB 2; Length 1457;			
Matches 1384; Conservative 37; Mismatches 36; Indels 33; Gaps 1;			
QY	1	MAGNSILVPIVLWGRKAPTHCISAVLLTDDGATITGCHDGOICLWDLSEVLQINPRALL	60
Db	1	MAGNSLVLPVILWGRKAPTHCISILLTDDGTTITGCHDGOICLWDLSEELVPRALL	60
QY	61	FIGHTASITCLSKACASSDKQYIVSASEGEMCLWDSGRCIEFTKLACTHTGTIOFYQFS	120
Db	61	FIGHTAAITCLSKACASGDQYTVSASANGEMCLWVDNGRCIEFTKLACTHTGTIOFYQFS	120
QY	121	VGNQREGRLCHGHVPEILVVDATSLVLYSLVSKI SPDWISSMSIIRSHRTQEDTVVAL	180
Db	121	VGNQREGRLCHGHVPEILVVDATSLVLYSLVSKI SPDWISSMSIIRSHRTQEDTVVAL	180
QY	181	SVTGILKWIIVTSEISDMQDTEPIFEESKPIYCONCOSISFCFTQRSLLVVCSEYWRV	240
Db	181	SVTGILKWIIVTSEISGLQDTEPIFEESKPIYCONCOSLSFCFTQRSLLVVCSEYWRV	240
QY	241	FDAGDYSLLCSGSPENGQTWCGDFVSDKVI IWTENGQSYIYKLPASCLPASDSFRSDV	300
Db	241	FDAGDYSLLCSGSPSDGQTWCGDFVSADKVI IWTENGQSYIYKLPASCLPASDSFRSDV	300
QY	301	GRAVENLIPVQHILLDRKDELLICPPVTRFPYCREYFHKLLIQGSSGRLNTWNISD	360
Db	301	GRAVENLIPVQHSHLLQKDELVI CPPVTRFPYCKEYHLHLLIQGSSGRLS IWN IAD	360
QY	361	TADKQGSREGLAMTTSISLOEAPDKLPCPAGI I DQLSVI PMSNEPLKVTASVYI PAHGR	420
Db	361	IADKQANEGLLKTTTCISLODADFKLKPCPAGI I DQLSVI PMSNEPLKVTASVYI PAHGR	420
QY	421	LYCGBEDGSIIVPATQTAIVQLLQGEHMLRRGPPHRTLGRHKNKVTCLLYPHQVSARY	480
Db	421	LYCGBEDGSI IIVPATQTAIVQLLQGEHMLRRGPPHRTLGRHKNKVTCLLYPHQVSARY	480
QY	481	DORYLISGGVDFSVI IWDIFSGEMKHI FCVHGGEITOLLVPPENCARSARVOHCICSVASDH	540
Db	481	DORYLISGGVDFSVI IWDIFSGEMKHI FCVHGGEITQLLLVPPENCARSARVOHCICSVASDH	540
QY	541	SVGLLSLEKKICIMLASRHLFPQVWKRPSPDYLVVGCSDGSVYVWMDTGALDRCVMG	600
Db	541	SVGLLSLEKKICIMLASRHLFPQVWKRPSPDYLVVGCSDGSVYVWMDTGALDRCVMG	600
QY	601	ITAVEILNACDEAVPAADVSLSHPAVNLQAMTRRSALAKKMAHHKLTQTLATNLLASEA	660
Db	601	ITAVEILNACDEAVPAADVSLSHPAVNLQAMTRRSALAKKMAHHKLTQTLATNLLASEA	660
QY	661	SDGNLPKYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIOLLTEASRPNTALISPENL	720
Db	661	SDGNLPKYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIOLLTEASRPNTALISPENL	720
QY	721	QKAGSSDKGSGFLTGKRAAVLFQOVKETI KENI KEHLIDDEEDEEIMRQREESDPEY	780
Db	721	QKAGSSDKGSGFLTGKRAAVLFQOVKETI KENI KEHLIDDEEDEEIMRQREESDPEY	780
QY	781	RSSKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKLPCHTQVSGLLSR	840
Db	781	RASKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKLPCHTQVSGLLSR	840
QY	841	GGHMSLMLPGYNOPACKLSHGKTEVGRKLPASBGVKGTYGVSRVTTQHLHSIISLANT	900
Db	841	GGHMSLMLPGYNQAAGKLHAKAEVGRKLPAAEGVGKGYTVSRVTTQHLHSIISLANT	900
QY	901	LMSMTNATFIGDHMKKGTPRPPRSTPDLISKARGSPPTSSNIVQGOIKQVAAPVVSARS	960
Db	901	LMSMTNATFIGDHMKKGTPRPPRSTPDLISKARDSPPASSNIVQGOIKQ	949
QY	961	ADHSGSDPPSAPALHTCFLVNEGWSQLAAMHCVMLPDLGLGDKFRPPLLEMLARRWQDRC	1020

Db 950 -----GHSQALAMHCVMPLDLLGLGKFRPPLLEMLARRWQDR 987

Qy 1021 LEVREAAQALLAEIRIEQAGRKEADAWAPYLQYIDHVISPGVTSEAAQTITAPDA 1080

Db 988 LEVREAAQALLAEIRIEQAGRKETITDWPYLPQYMDHVISPGVTAEAMQWAAAPDA 1047

Qy 1081 SGPEAKVQEEHDLVDDDDITTCGLSSVPQMKKISTSEERKQATAIVLLGVIGAEFGAE 1140

Db 1048 SGPEAKVQEEHDLVDDDDITTCGLSSVPQMKKISTSEERKQATAIVLLGVIGAEFGAE 1107

Qy 1141 IEPKLLTRPRSSQIPGEGCLTSGGNSYLARHTCKALTPLLOPPSPKLPHPHSTIRRT 1200

Db 1108 IEPKLLTRPRSSQIPGEGCLTSGGNSYLARHTCKALTPLLOPPSPKLPHPHSTIRRT 1167

Qy 1201 AIDLIGRFTWPEYMDVSAYLMLGELCADAERKQALNITMGLPLSPAADSARSARHALS 1260

Db 1168 AIDLIGRFTWPEYMDVSAYLMLGELCADAERKQALNITMGLPLSPAADSARSARHALS 1227

Qy 1261 LIATARPAPFITTAKVHRHTALANTQSOONHTTTLARAKGEILRVIEILIEKMPDT 1320

Db 1228 LIATARPAPFITTAKVHRHTALANTQSOOSHTTTLARAKGEILRVIEILIEKMPDT 1287

Qy 1321 VVDLLVEVMDIIMYCLSGSLVKKGLQECFPAICRFYMWVSYYERNHRIAVGARHGSVALY 1380

Db 1288 VVDLLVEVMDIIMYCLSGSLVKKGLQECFPAICRFYMWVSYYERNHRIAVGARHGSVALY 1347

Qy 1381 DIRTKGQTIHGHKGPITAVAFADPGRYLATYSNTDISHISFWQNTSLLGSLGMLNSAPQ 1440

Db 1348 DIRTKGQTIHGHKGPITAVAFADPGRYLATYSNTDISHISFWQNTSLLGSLGMLNSAPQ 1407

Qy 1441 LRCIKTYQVPPVPASPQSHNALKARLWTSNRNVILMAHDGKEHRFMV 1490

Db 1408 LRCIKTYQVPPVPASPQSHNALKARLWTSNRNVILMAHDGKEHRFMV 1457

RESULT 7

WDR7 HUMAN STANDARD; PRT; 1160 AA.

AC Q9Y4E6;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE WD-repeat protein 7 (Fragment).

GN Name=WDR7; Synonyms=KIAA0541;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=98290545; PubMed=9628581;

RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,

RA Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. IX.

RT The complete sequences of 100 new cDNA clones from brain which can

RT code for large proteins in vitro.";

RL DNA Res. 5:31-39(1998)

CC -1- SIMILARITY: Contains at least 6 WD repeats.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; AB011113; BAA25467.1; -.

DR PIR; T00272; T00272.

DR Genew; HGNC:13490; WDR7.

DR InterPro; IPR001680; WD40.

DR InterPro; IPR011046; WD40_like.

Pfam; PF00400; WD40; 4.

DR PRINTS; PR00320; GPROTEINBRPT.

DR SMART; SM00320; WD40; 4.

DR PROSITE; PS00678; WD_REPEATS_1; 1.

DR PROSITE; PS00082; WD_REPEATS_2; 3.

DR PROSITE; PS0294; WD_REPEATS_REGION; 2.

KW Repeat; WD repeat.

FT NON TER 1 1

FT REPEAT 1 36 WD 1.

FT REPEAT 74 113 WD 2.

FT REPEAT 132 177 WD 3.

FT REPEAT 228 267 WD 4.

FT REPEAT 1021 1060 WD 5.

FT REPEAT 1062 1102 WD 6.

SQ SEQUENCE 1160 AA; 127452 MW; A3784B351832484D CRC64;

Query Match 77.58; Score 6023; DB 1; Length 1160;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 331 RFFYGCREFYFKLLIQDSSGRNLNWNISDTADKQSEEGLAMTTSISLQEAFLKLP 390

Db 1 RFFYGCREFYFKLLIQDSSGRNLNWNISDTADKQSEEGLAMTTSISLQEAFLKLP 60

Qy 391 AGIIDQLSVIPNSNEPLKVTASVYIPAHGRVLCREDGSIIVIPATQTAVQLLQGEHML 450

Db 61 AGIIDQLSVIPNSNEPLKVTASVYIPAHGRVLCREDGSIIVIPATQTAVQLLQGEHML 120

Qy 451 RRGWPPHRTLRGHRNKVTCLLYPHQVSARYDQRYLISGGVDFSVIIWDIFSGEMKHIFCV 510

Db 121 RRGWPPHRTLRGHRNKVTCLLYPHQVSARYDQRYLISGGVDFSVIIWDIFSGEMKHIFCV 180

Qy 511 HGGEITQLLVPPENCARSARVQHCCI CSVASDSHVGILLSREKKCMILASRHLFPPIQVIKWRP 570

Db 181 HGGEITQLLVPPENCARSARVQHCCI CSVASDSHVGILLSREKKCMILASRHLFPPIQVIKWRP 240

Qy 571 SDDYLVVGCSDGSVYVQMDTGALDRCVMGITAVEILNACDEAVPAADVLSLSHPAVNLKQ 630

Db 241 SDDYLVVGCSDGSVYVQMDTGALDRCVMGITAVEILNACDEAVPAADVLSLSHPAVNLKQ 300

Qy 631 AMTRSLAALKNNWAHKLQTLATNLLASEADSKGNLPKYSHNSLMVQAIKTNLTDPDIHV 690

Db 301 AMTRSLAALKNNWAHKLQTLATNLLASEADSKGNLPKYSHNSLMVQAIKTNLTDPDIHV 360

Qy 691 LFFDVEALIIQLLTERASRENTALISPENLQKASGSDGSGFLTGKRAAVLFOQVKETI 750

Db 361 LFFDVEALIIQLLTERASRENTALISPENLQKASGSDGSGFLTGKRAAVLFOQVKETI 420

Qy 751 KENIKEHLDDDEEDEEIMQRREESDPYRSSKSKPLTLLEYNLTMDTAKLFMSCLHAW 810

Db 421 KENIKEHLDDDEEDEEIMQRREESDPYRSSKSKPLTLLEYNLTMDTAKLFMSCLHAW 480

Qy 811 GLNEVLDEVCLDRGLMKLPHCTVSFGLLSRGHMSLMPCYNQPACKLSHGKTEVGRKLP 870

Db 481 GLNEVLDEVCLDRGLMKLPHCTVSFGLLSRGHMSLMPCYNQPACKLSHGKTEVGRKLP 540

Qy 871 ASSEGVGKGTGVSRVATTQHLLSIIISLANTLMSMTNATFIGDHMKGKPTPRPSTPDL 930

Db 541 ASSEGVGKGTGVSRVATTQHLLSIIISLANTLMSMTNATFIGDHMKGKPTPRPSTPDL 600

Qy 931 KARGSPPTSSNIYVQGIKQVAPVWSARSADHSGSDPPSPALHTCFLVNEGWSQLAAM 990

Db 601 KARGSPPTSSNIYVQGIKQVAPVWSARSADHSGSDPPSPALHTCFLVNEGWSQLAAM 660

Qy 991 HCWMLPDLGLDKFRPPLLEMLARRWQDRCLVREAAQALLAEELRRIEQAGRKEAIDAW 1050

Db 661 HCWMLPDLGLDKFRPPLLEMLARRWQDRCLVREAAQALLAEELRRIEQAGRKEAIDAW 720

Qy 1051 APYLPQYIDHVISPGVTSEAAQTITAPDASGPEAKVQEEHDLVDDDDITTCGLSSVPQM 1110

Db 721 APYLPQYIDHVISPGVTSEAAQTITAPDASGPEAKVQEEHDLVDDDDITTCGLSSVPQM 780

Qy 1111 KKISTSYEERRKQATAIVLLGVIGAEFGAIEBPPKLLTRPRSSSQIPEGFGLTSGGSNYS 1170

```
Db 781 KKTSTSEERKQATAIVLLGVIGAEFGAIEPPKLLTRPRSSQIPEGFGLTSGGSNY 840
Qy 1171 LARHTKALTFLLQPPSPKLPKPHSTIRRTAIDLIGRGFTWBPYMDVSAVLMGLLELCA 1230
Db 841 LARHTKALTFLLQPPSPKLPKPHSTIRRTAIDLIGRGFTWBPYMDVSAVLMGLLELCA 900
Qy 1231 DAEKOLANTWGLPLSPAADSARSARHALSIATARPAPITTTIAKEVHRHTALAANTOS 1290
Db 901 DAEKOLANTWGLPLSPAADSARSARHALSIATARPAPITTTIAKEVHRHTALAANTOS 960
Qy 1291 QQNHHITTLARAKEILRVIEIIEKMPDVTVDLLVEVMDIIMYCLEGSLVKKKGLQECF 1350
Db 961 QQNHHITTLARAKEILRVIEIIEKMPDVTVDLLVEVMDIIMYCLEGSLVKKKGLQECF 1020
Qy 1351 PAICRFYMSYYERNHRIAGARHGSVALYDITRGKQCTTHGHKGPITAVAPADGRYLA 1410
Db 1021 PAICRFYMSYYERNHRIAGARHGSVALYDITRGKQCTTHGHKGPITAVAPADGRYLA 1080
Qy 1411 TYSNTSHISFWQNTSLGSGMNSAPOLRCIKTYQVPPVQSPGSHNALKLARLIW 1470
Db 1081 TYSNTSHISFWQNTSLGSGMNSAPOLRCIKTYQVPPVQSPGSHNALKLARLIW 1140
Qy 1471 TSNRNVLMAHDGKEHRFMV 1490
Db 1141 TSNRNVLMAHDGKEHRFMV 1160

RESULT 8
Q80TY3 PRELIMINARY; PRT; 1166 AA.
AC Q80TV3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE MKIAA0541 protein (Fragment).
DE Name=Wdr7; Synonyms=mkIAA0541;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;
RA Okazaki N., Kikuno R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT ii. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48(2003).
CC -1- SIMILARITY: Contains 4 WD repeats.
DR EMBL; AKI22304; BAC65586.1; -.
DR MGD; MGI:1860197; Wdr7.
DR InterPro; IPR001680; WD40.
DR IntraPro; IPR011046; WD40_like.
DR Pfam; PF00400; WD40; 4.
DR PRINTS; PR00320; GPROTEINRPT.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
DR PROSITE; PS50082; WD_REPEATS_2; 3.
DR PROSITE; PS50294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
FT NON TER 1
SQ SEQUENCE 1166 AA; 128360 MW; 0D4B66A78F262D4D CRC64;

Query Match 74.48; Score 5784.5; DB 2; Length 1166;
Best Local Similarity 92.98; Pred. No. 0;
Matches 1114; Conservative 27; Mismatches 33; Gaps 1;

Qy 292 ASDSFRSDVGKAVENLIPPVQHSLLDQDKELVICPPVTRFFYCKEYLHKLIIQDSSG 351
|||||
```

```
Db 1 ASDSFRSDVGKAVENLIPPVQHSLLDQDKELVICPPVTRFFYCKEYLHKLIIQDSSG 60
Qy 352 RLNIWNIISTADKQSGEEGLAMTTTISLOEAFDKLNPCAGIIDQLSVTPNSNEPKVTA 411
Db 61 RLNIWNIADIAEQEADEGLKMTTISLOEAFDKLPCAGIIDQLSVTPNSNEPKVTA 120
Qy 412 SVYIPAGRLVCGREGDSIVIVPATQTAIVQLLQGEHMLRRGPPHRTLRGHENKVTCLL 471
Db 121 SVYIPAGRLVCGREGDSIIIVPATQTAIVQLLQGEHMLRRGPPHRTLRGHENKVTCLL 180
Qy 472 YPHQVSARYDQRYLISGGVDFSVIIWIDIFSGEMKHFVCHGGEITOLLVPPENC SARVOH 531
Db 181 YPHQVSARYDQRYLISGGVDFSVIIWIDIFSGEMKHFVCHGGEITOLLVPPENC SARVOH 240
Qy 532 CTCVSASDSHVGILSLREKKCMLASRHLFPQIVIKWRPSDDYLVVGCSDGSGVYVWMDT 591
Db 241 CTCVSASDSHVGILSLREKKCMLASRHLFPQIVIKWRPSDDYLVVGCSDGSGVYVWMDT 300
Qy 592 GALDRVCVMGITAVEIILNACDEAVPAVDSLSHPAVNLKQAMTRRSALAKNMAHKKLQTL 651
Db 301 GALDRVCAMGITAVEIILNACDEAVPAVDSLSHPAVNLKQAMTRRSALAKNMAHKKLQTL 360
Qy 652 ATNLLASEASDKGNLPKYSHNSLMVQAIKTNLTDPIHVLFPDVEALIIQLITEASRPN 711
Db 361 ATNLLASEASDKGNLPKYSHNSLMVQAIKTNLTDPIHVLFPDVEALIIQLITEASRPN 420
Qy 712 TALISPENLOKAGSGSDKGSFLTGKRAAVLFOOVKETIKENIKHLLDDEEBEIMRQ 771
Db 421 TALISPENLOKAGSGSDKGSFLTGKRAAVLFOOVKETIKENIKHLLDDEEBEIMRQ 480
Qy 772 RREESDPEYRSKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHC 831
Db 481 SRESDSPEYRASKSKPLTLLEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHC 540
Qy 832 TVSFGLLSRGHMSLMPLPGYNOPACKLSHGKTEVGRKLPASEGVGKGTGVSAVTTQHL 891
Db 541 TVSFGLLSRGHMSLMPLPGYNQAGKLLHAKAEVGRKLPAAEGVGKGTGVSAVTTQHL 600
Qy 892 LSTIISLANTLMSMTNATFIDGMKGPTRPPRSTPDLKARGSPPTSSNIVOGQIKOVA 951
Db 601 LSTIISLANTLMSMTNATFIDGMKGPTRPPRSTPDLKARGSPPTSSNIVOGQIKOVA 658
Qy 952 APVVSARSADHSGSDPPSAPALHTCFLVNEGWSLAAAMHCVMLPDLGLGDKFRPPLLEM 1011
Db 659 -----GWSQLAAMHCVMLPDLGLGLERPRPPLLEM 687
Qy 1012 LARRWQDRCLVEVREAAQALLLAELRRIEAGRKEAIDAWAPYLPQIVDHVISPVTSEAA 1071
Db 688 LARRWQDRCLVEVREAAQALLLAELRRIEAGRKEAIDAWAPYLPQIVDHVISPVTSEAA 747
Qy 1072 QTITTPADASGPEAKVQEEHDLVDDITTCGLSSVPQMKCISTSYEERRKQATAIVLLG 1131
Db 748 QTMAAADASGPEAKVQEEHDLVDDITTCGLSSVPQMKCISTSYEERRKQATAIVLLG 807
Qy 1132 VIGAEGFAETPEPKLLTRPRSSQIPEGFGLTSGGSNYSLARHTCKALTALLQPPSPKL 1191
Db 808 VIGAEGFAETPEPKLLTRPRSSQIPEGFGLTSGGSNYSLARHTCKALTALLQPPSPKL 867
Qy 1192 PPHSTIRRTAIDLIGRGFTWBPYMDVSAVLMGLLELCADAEKOLANI TWGLSPADS 1251
Db 868 PPHSTIRRTAIDLIGRGFTWBPYMDVSAVLMGLLELCADAEKOLANI TWGLSPADS 927
Qy 1252 ARSARHALSIATARPAPITTTIAKEVHRHTALAANTQSOONHMTTLARAKEILRVIE 1311
Db 928 ARSARHALSIATARPAPITTTIAKEVHRHTALAANTQSOOSHTTTLARAKEILRVIE 987
Qy 1312 ILIEKMPDVTVDLLVEVMDIIMYCLEGSLVKKKGLQECFPAICRFYMSYYERNHRIAVG 1371
Db 988 ILIEKMPDVTVDLLVEVMDIIMYCLEGSLVKKKGLQECFPAICRFYMSYYERNHRIAVG 1047
Qy 1372 ARHGSVALYDITRGKQCTTHGHKGPITAVAPADGRYLATYSNTDSHISFWQNTSLGSG 1431
Db 1048 ARHGSVALYDITRGKQCTTHGHKGPITAVAPADGRYLATYSNTDSHISFWQNTSLGSG 1107
```

QY 1432 IGMNSAPQRCIKTYVPPVQSPASPGSHALKLARLIWTSNRNVILMAHDGKEHRPMV 1490
DB 1108 IGMNSAPQRCIKTYVPPVQSPASPGSHALKLARLIWTSNRNVILMAHDGKEHRPMV 1166

RESULT 9
Q6GPQ7 Q6GPQ7 PRELIMINARY; PRT; 956 AA.

AC Q6GPQ7;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE MGC82685 protein.

GN Name=MGC82685;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE=Ovary;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diatchenko L., Marusina K., Farmer A.S., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uudin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gumaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.

RC TISSUE=Ovary;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.B., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
initiative.";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP SEQUENCE FROM N.A.

RC TISSUE=Ovary;
RX Klein S., Gerhard D.S.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 5 WD repeats.
DR EMBL: BC073055; A073055.1; -.
DR InterPro: IPR01680; WD40.
DR Pfam: PF00400; WD40; 5.
DR PRINTS: PR00320; GPROTEINERT.
DR SMART: SM00320; WD40; 5.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.
DR PROSITE: PS00682; WD_REPEATS_2; 3.
DR PROSITE: PS0294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 956 AA; 105601 MW; 23A89D893460F923 CRC64;

Query Match 55.7%; Score 4331; DB 2; Length 956;
Best Local Similarity 85.1%; Pred. No. 1.1e-280;

Matches 808; Conservative 73; Mismatches 68; Indels 0; Gaps 0;

QY 1 MAGNSLVLPVLMGRKAPTHCISAVLLTDDGATVITGCHDQICLMDLSVELQINPRALL 60
DB 1 MAGNSLVLPVLMGRKAPTHCISAVLLTDDGATVITGCHDQICLMDLSPELEITPRALL 60

QY 61 FGTASITCLSKACASSDKOYIVSASESGEMCLMDYDGRCEIETFKLACTHTGTIOYOPS 120
DB 61 FGTASITCLSKACASSDKOYIVSASESGEMCLMDYDGRCEIETFKLACTHTGTIOYOPS 120

QY 121 VGNQREGRLCHGHYPEILVWDATSLVLYSLVSKISPDWISSMSIIRSHRTQEDTVVAL 180
DB 121 VGNQREGRLCHGHYPEILVWDATSLVLYSLVSKISPDWISSMSIIRSHRTQEDTVVAL 180

QY 181 SVTGILKWIWTVTSEISDMQDTEPIFEESKPIYQNCQSFCAFTQSLVVCVKYRW 240
DB 181 SVTGILKWIWTVTSEISDMQDTEPIFEESKPIYQNCQSFCAFTQSLVVCVKYRW 240

QY 241 FDAGDYSLICSGPSENGQTWTGDFVSSDKVLIWTENGQSYIYKLPASCLPASDSFRSDV 300
DB 241 LDAGDFSLICSGPSENGQTWTGDFVSSDKVLIWTENGQSYIYKLPASCLPASDSFRSDV 300

QY 301 GRAVENLIPPVOHILADKXKELLICPPVTRPFYGCVEYFHKLLIQDSSGRININWISD 360
DB 301 GRAVENLIPPVOHILADKXKELLICPPVTRPFYGCVEYFHKLLIQDSSGRININWISD 360

QY 361 TADKQSESEGLAMTTSISLQEAFDKLNPAGIIDLQSVIPNSNBEPLKVTASVYIPAHGR 420
DB 361 TADKQSESEGLAMTTSISLQEAFDKLNPAGIIDLQSVIPNSNBEPLKVTASVYIPAHGR 420

QY 421 LVCGREDGSIIVPATQTAIVQLQGEHMLRGWPPHRTLRGHRNKVTCCLLYPHQVSARY 480
DB 421 LVCGREDGSIIVPATQTAIVQLQGEHMLRGWPPHRTLRGHRNKVTCCLLYPHQVSARY 480

QY 481 DORYLISGGVDSVSVIWDIIFSGEMKHFVHGGEITQLLVPPENCARVOHCITCSVASDH 540
DB 481 DORYLISGGVDSVSVIWDIIFSGEMKHFVHGGEITQLLVPPENCARVOHCITCSVASDH 540

QY 541 SVGLLSLRKCKIMLASRHLFPQVTKWRPSDDYLVVGCSDGVSYYVQMDTGALDRCVMG 600
DB 541 SVGLLSLRKCKIMLASRHLFPQVTKWRPSDDYLVVGCSDGVSYYVQMDTGALDRCVMG 600

QY 601 ITAVEILNACDEAPVPAVDSLSHPAVNLQAMTRRSIALKMAHKLQTLATNLASEA 660
DB 601 ITAVEILNACDEAPVPAVDSLSHPAVNLQAMTRRSIALKMAHKLQTLATNLASEA 660

QY 661 SDKGNLPKYSHNSLMVQAIKTNTDPIHVLFPDVEALIIQLLTERASRNTALISPENL 720
DB 661 SDKGNLPKYSHNSLMVQAIKTNTDPIHVLFPDVEALIIQLLTERASRNTALISPENL 720

QY 721 QKASGSSDKGSGFLTGKRAAVLFOQVKETIKENIKEHLDDDEEDEEIMRQREESDPEY 780
DB 721 QKASGSSDKGSGFLTGKRAAVLFOQVKETIKENIKEHLDDDEEDEEIMRQREESDPEY 780

QY 781 RSKSKPLTLLEYNLTMTAKLFMSCLHAWGLNEVLDEVLDRGLMKPHCTVSGFLLSR 840
DB 781 RSKSKPLTLLEYNLTMTAKLFMSCLHAWGLNEVLDEVLDRGLMKPHCTVSGFLLSR 840

QY 841 GGHMSLMYPGYNQPACKLSHGKTEVGRKLPASEGVCKGTGYSVRAVTTQHLISILANT 900
DB 841 GGHMSLMYPGYNQPACKLSHGKTEVGRKLPASEGVCKGTGYSVRAVTTQHLISILANT 900

QY 901 LMSMTNATFIDHMKKGPTRPPRSPDLSKARGSPPTSSNIVQOIKQ 949
DB 901 LMSMTNATFIDHMKKGPTRPPRSPDLSKARGSPPTSSNIVQOIKQ 949

RESULT 10
Q86VP2 Q86VP2 PRELIMINARY; PRT; 614 AA.
AC Q86VP2;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)


```
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE WDR7 protein (Fragment).
GN Name=WDR7;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.L., Wang J., Heieh F.,
RA Diatchenko L., Marudina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalusz D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RA Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 WD repeat.
DR EMBL: BC050352; AH50352.1; -.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 1.
DR SMART: SM00320; WD40; 1.
DR PROSITE: PS50082; WD_REPEATS_2; 1.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW WD repeat.
FT NON_TER
SQ
Query Match 40.8%; Score 3170; DB 2; Length 614;
Best Local Similarity 100.0%; Pred. No. 2.9e-203;
Matches 614; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 877 KGTGYGVSRAVTTQHLISIIISLANTLMSMTNATFIGDHMKKGTPRPPRSTPDLKARGSP 936
DB 1 KGTGYGVSRAVTTQHLISIIISLANTLMSMTNATFIGDHMKKGTPRPPRSTPDLKARGSP 60
QY 937 PTSSNVVQGIKQVAPVVSARGSDADSHSGSDPPSPALHTCFVLNCGWSQLAAHWCMLP 996
DB 61 PTSSNVVQGIKQVAPVVSARGSDADSHSGSDPPSPALHTCFVLNCGWSQLAAHWCMLP 120
QY 997 DLGLDKFRPPLLEMLARRWQDCLVREAAQALLAEIRIIOAGKKAIDAWAYLPQ 1056
DB 121 DLGLDKFRPPLLEMLARRWQDCLVREAAQALLAEIRIIOAGKKAIDAWAYLPQ 180
QY 1057 YIDHVISPGVTSEAAQTITTPADSGPEAKVQEEHDLVDDDTTTCGLSVPQMKKISTS 1116
DB 181 YIDHVISPGVTSEAAQTITTPADSGPEAKVQEEHDLVDDDTTTCGLSVPQMKKISTS 240
QY 1117 YEERRKQATVILLVIGAFGEAIEPPKLLTTPRSSQITPEFGLTSGGSNYSLARHTC 1176
DB 241 YEERRKQATVILLVIGAFGEAIEPPKLLTTPRSSQITPEFGLTSGGSNYSLARHTC 300
QY 1177 KALTFTLLQPPSPKLPHPHSTIRRTAIDLIGRGFTVWPEYMDVSAVLMLGLELCAEAKQL 1236
```



```
Db 1032 IREAAQIILLGELTRMGKGRKQLVESQAQYPLLYTH--TEPIVGAQQQALALISQPASGG 1089
Qy 1079 -----DASGPEAKV-----QBEHDLVDDITTTGCLSVPMQ 1110
Db 1090 AGSGSGGNGGVGVSGGGAGSGGPGGVPGGDAHODEDEYEBEEI- 1139
Qy 1111 KKTSTSEERKQATAIVLLGVGAEGABE--EPPK-----LLTRPRS 1152
Db 1140 KPESSISELKKTKTAVILLGVGAEGQDISESPNHRGISTMATGANLTSGVAGERR 1199
Qy 1153 SSOIPGFGLTSGGSNSYLARHTCKALTFLLLQPPSPKLPPLPHSTIRRTAIDLGRGFTVM 1212
Db 1200 KSSVVEGFGIAN-----NLARLTSMALHLLYAPPSPKLPQYTPLRRAALDILLGRGFTVM 1254
Qy 1213 EPMDVSAVMGLLELCADAEKOLANITWGLPLSPAADSARSARHLSLIATARPFAFIT 1272
Db 1255 EPYLDVSKVLLGLLEISCEG-KAVPNLNYKLPITPOADACTARHALRLIATARPAFIT 1313
Qy 1273 TIAKEVHRHTALAAQTQQNMHT-TTLARAKGEILRLVTEILIEKPTDVVDLLVEVMDI 1331
Db 1314 TMAREVARVNTMOQNAQSINTPLTQSVLHKAKGEILQCVEMLIDRKQSEIAGLLVEVMDI 1373
Qy 1332 IMVCLGSLVKKGLQECFPAICRFYMYSYERNHRIAVGARHGSVALYDIRTKCQTIH 1391
Db 1374 ALHCVDGNELKNRGLAELCPAICKFNQISHCAQTRRIAVGANSGLAIYELRQNKQMP 1433
Qy 1392 GHGKPTTAVAPADPGRYLATYSNTDHSISPMQNTSLLSIGMNSAPQLRCITQVPP 1451
Db 1434 AHTHPTISLAFSPDGKYLVSYSACENRLSPWQSTGMFG-LGQ----SOTRCTKGYSTAP 1488
Qy 1452 VQPASGSHNALKLARLITNSNRNVLIMAHGDKERPMV 1490
Db 1489 IPDVS---RLNPMRLAKLWLNNTVTMLMLADGSETRFN 1525

RESULT 12
Q7Q3S0 PRELIMINARY; PRT; 1473 AA.
AC Q7Q3S0;
DT 01-MAR-2004 (T-EMBLrel. 26, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE EbiP7067 (Fragment).
GN Name=ebic7067; ORFNames=ENSANGG00000005326;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: Contains 5 WD repeats.
CC -I- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AAAB01008964; EAA12375.1; -.
DR InterPro; IPR001680; WD40.
DR InterPro; IPR011046; WD40_like.
DR Pfam; PF00400; WD40; 5.
DR PRINTS; PR00320; GPROTEINBRPT.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_2.
DR PROSITE; PSS0082; WD_REPEATS_2; 2.
DR PROSITE; PSS0294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
FT NON_TER 1473 1473
SQ SEQUENCE 1473 AA; 163970 MW; 5074DBD5F82A1AA7 CRC64;

Query Match 40.0%; Score 3107.5; DB 2; Length 1473;
Best Local Similarity 42.1%; Pred. No. 1.8e-198;
Matches 659; Conservative 285; Mismatches 452; Indels 171; Gaps 35;
```

```
Qy 1 MAGNSLVLPVLVGRKAPTHCISAVLLTDDGATVGTGCHDGOICLWDLSE-LQINPRAL 59
Db 1 MANTNLVVPVLVGNPTPHCVSSVPLSRDQKILATGCGYDQICLWQVDPDTLQMTPRCL 60
Qy 60 LFGHTASITCLSKACASSDKQYIVSASEGEMCLMDVSDGRCEIEFTKLACTHTGIFYQF 119
Db 61 LVGHTAPVLCRTASIIQDNPLVSSSENGEMCTDLVDGKCTESIKMPQVHTNIQAYH- 119
Qy 120 SVGNQBEGLLCHGYPEILVVDATSLVLSVSKI SPDWISSMSIISHSRTQEDTVA 179
Db 120 -MANCDOIIFLCNGYAEIMVDPFSLVFLCLSSKVNPDWISALHLPKSRKDDVILA 178
Qy 180 LSVTGILKWIIVTSETSDMODTEPIPEESKPIYCONCOSISFCAPTORSLVSVCSKYR 239
Db 179 ITTGTGVKWTLLG--NENKSEPIYENESKQIRCLNALTMCNCSQNRQTVLIVCAKYQ 236
Qy 240 VFDAGYSLLCSGSPSNGQTGGDPVSSDKVIITWENGOSYIYKLPASCLPASDSFRS- 298
Db 237 IYDAGDFTVLCSVISPSGERWGGDFLSNDRVLMSDEGKYLFRLPANSI PNKDHPG 296
Qy 299 DYGKAVENLIPVQHILLDRKDELLICPVTRFFYGCRE-YPHKLLIOGSSGRLLIWN 357
Db 297 SVGKDT-----PFLYCILSQTGKPLSCPPTMKLVTSNRSGTKQKYLGRDSEGYNIWA 351
Qy 358 ISDTADKQGE-----EGLAWTTSISLQAFDKLNPAGIIDQLSVIPNSNEPLKV 409
Db 352 VPDISEQIKQIQTKQHPASLESTICTSLLEAWKGNPLPVGILDQL--YQNSEPWIKL 409
Qy 410 TASVYIPAHGRIVCGREDGSIIVIPATQTAIVOLIGEHMLRRGWPHHTLGRHNKVT 469
Db 410 TASIYLPQOSRLVVGREDGSIIVPATQTVMLHLHGNHQFSDWPHQILANGHSRVNC 469
Qy 470 LLYPHQVSARYQRYLISGVDFSVIWDI FGEKMHIFCVHGEITOLLVPPENCASV 529
Db 470 LLCPSLAHSRYDKSHLVSGGVDFAVCLMDLYSGSLIHRFCVHAGEITOLLVPPPTCSPRI 529
Qy 530 QHCICSVASDHSVGLLSLREKICIMLASRHPPIQVIKWRPDDYLVVCGSDGVVWQM 589
Db 530 LKICISVADSHSVTLISLQERKCVLASRHLFPVISIKWRPLDDFLIVGSDGVVWQM 589
Qy 590 DTGALDRVCVMGITAVELNACDEAVPAVDS-----LSHPAVNLKQAMTRSLAALK 641
Db 590 ETGHLDRVLHGLAEVLNACDENSGETSGSTSEMGLANPAVHFFRGLRHNINAIR 649
Qy 642 NMAH---HKLQYLATNLLASEADKGN-LPKYSHNSLAWQAKTNLTDPDIHVLFPDVEA 697
Db 650 HATQGIHQLOQ-----LHANHQHGFLLMKNRSSPLIIQGLRTPKDAESHILFFDIEG 704
Qy 698 LIIQLLTER--ASRPNTALISPENLOKASGSDKGSFLTGKRAAVLFOQVK-----E 748
Db 705 LIFELHAEYVAMSPT--LEAEGLLIPTGTDHSD--AGKKISDFGRVKNKAGDMEX 760
Qy 749 TIKENIKEHLDDDEBEDEIMQRRE-----ESDPEYRSKSKPLTLLEYNLTW 797
Db 761 ILKEDKHGILAKMKEGAENVQTKVQAKLESVVXNVNGETGKDSDDITTKIAPMEATHVM 820
Qy 798 DTAKLPMSCLHAGLNEVLDECLDGLMKLKHCTVSGFLSRGGHMSLWLPY-NOPAC 856
Db 821 EVAQALLSLIHWGLDPLDKVCETQGLGLRPMIPVSGVLGKGGVMSLLPTWQNNIVI 880
Qy 857 KLSHGKTEVGRKLP--ASEGVGK-----GTGVSVRAVTTQHLLSIISLANTLMSMTNAT 908
Db 881 NSTADELKTAMTPEQFRQEGTLKPTARLHWELSTYLSNHLANVANSNTLMSNMAT 940
Qy 909 FIGDHMKGTPRPPRSTPDLSKAR-----GSPPTSSN-----IVQGQIKQVAPV 954
Db 941 FI-----PEQERARKLHRQSTRATWSNEEBEQEAFTQOQAQIKQ----- 979
Qy 955 VSARSADHSGSDPPSPAPALHTCFLVNEGWSOLAAMHCVMKP---DLIGDKFRPPLLEM 1011
Db 980 -----GWSLSTHCHFLPDKIKDALDANNFRQVEM 1011
Qy 1012 LARRWQDRCLEVREAAQALLAEALRRIEQAGRKEAIDAWAPYLPQY--IDHVI--SPGVT 1067
```

```
Db 1012 MARWQHCHLEIRAAQOLLEGRMGKGRKQLVESWAGYLPMTHTHTPIVQOAPATG 1071
Qy 1068 SEAAQTITAPDAS-GPEAKVQEEHDLVDDITTCGLSSVPQMKKISTSYEBRRKQATA 1126
Db 1072 QSNAGSPTSPTGQGLEHEESEBEV-----VRKPSSLAELKRRQTTA 1117
Qy 1127 IVLGVIGAFGAIEIPP--KLTRPRSSQIPGEGPGLTSGGNSYSLARHTCKALFLLL 1184
Db 1118 VLLGVIGAFGGQDISATDGKRSNENRRKSSVVEGFI-----GNNNLARSTSMALTHLL 1173
Qy 1185 QPSPKLPHTTTRRTAIDILIGFTVWEPMVDSAVLMGLLELCADABKOLANITWGLP 1244
Db 1174 APATQKLPATPUPRAAIDILIGFTVWEPIIDVSKVLLGLECCDSNRLIPSINVKLP 1233
Qy 1245 LSPAADSARSARHALSIATARPAPFITTIAKEVHRHTALAANTQS-QQNMHTTTILARAK 1303
Db 1234 LTPQADACTARHALRIATARPAFITTWAREVARNTWQQAQIAVSPIITOSVLHRAK 1293
Qy 1304 GEILRVIEILIEKPTVDVLLVEWMDIMYCLEGLSVKKGLOECPPALCRFPMVSYE 1363
Db 1294 REILQCEMLIDKMQTEISNLLVEWMDITLHCVDSDGLKKNGLAEVSPLMCKFNQVSHCS 1353
Qy 1364 RNHRIAGARGHGSVALVDITGKCTIHKHGPITAVAPADGRYLATYNTDISHIFWQ 1423
Db 1354 ASRIIAGASNGHLAIYELKQNKQMPATKQVTALAPSPDGKFLVSYSCENRSLFWQ 1413
Qy 1424 MNTSLGSIOMLSAPOLRCIKTYQVPPQVOPASPGSHSNALKLARLWTSNRNVLMAHOG 1483
Db 1414 TSAGMFG-LGQ-----SQTRCIKGYSTAPIIDI--GRLNPMRLAKLIWINNRTVTLMLADG 1466
Qy 1484 KEHRFMV 1490
Db 1467 SETRFNV 1473

RESULT 13
ID O46092 PRELIMINARY; PRT; 1471 AA.
AC O46092;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE EG:86E4.3 protein.
GN Name:EG:86E4.3; ORFNames=CG17766;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferraz C., Vidal S., Brun C., Bucheton A., Demaille J.G.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 5 WD repeats.
DR EMBL; AL021086; CAAL5934.1; -.
DR FlyBase; FBgn0023510; CG17766.
DR InterPro; IPR01680; WD40.
DR InterPro; IPR011046; WD40_like.
DR Pfam; PF00400; WD40; 5.
DR PRINTS; PR00320; GPROTEINBRPT.
DR SMART; SM00320; WD40; 5.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_2.
DR PROSITE; PS00082; WD_REPEATS_2; 1.
DR PROSITE; PS0294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 1471 AA; 162496 MW; C468F629250DB2F2 CRC64;
```

Query Match

37.9%; Score 2944; DB 2; Length 1471;

```
Best Local Similarity 40.3%; Pred. No. 1.6e-187;
Matches 622; Conservative 294; Mismatches 447; Indels 182; Gaps 31;

Qy 54 INPRALLFGHTASITCLSKACASSDQYIVSASEGEMCLMDVSDGRCIEFTKLACTHTG 113
Db 1 MSPRCLLVGHSAPVLCIVRASLLPENFLVSSSENGEMCTWDLTDGKCEAVKLPQVHTQ 60
Qy 114 IQYQSVGNQREGRLLCHGYPEIILVVDATSLVLYSLVSKISPDWISMSIIRSHRTQ 173
Db 61 IQSYH--TANSEDVRLFCIGYAEIMVMDPFSLEVIVLSSKVKPDMWISAIHRLPMRRK 118
Qy 174 EDTVVALSVTGILKWIWVTSIDMDTPEIPEESKPIYQNCQSTISFCAPTQRSLLVV 233
Db 119 DDVLAITTTGVTKVWTLTG--NENKHAEPYENESKEIRCLNAINTWCCAQNRQTVLLV 176
Qy 234 CSKYRWVFDAGDYSLLCSGSPENGOTWGTGDFVSSDKVIIWTENGOSIYIKLPAJCLPAS 293
Db 177 CTYQWQIYDAGDFTVLCVAPARERWQSGDFITSDRVMLTDEGKLYLKLPAJCLPDI 236
Qy 294 DFRSDVGKAVENLIPVQVHILLDRKELLICPPVTRFPYGYCREYFHKLLIOGDSGRL 353
Db 237 KEFHS---KSVVRDAPLYLVYVQHAGDK-VLSCPPAMKLIQAGAGQH--LLRGDSSEGI 290
Qy 354 NTW-----NLSDTADKQSGEEGLAMTTSISLQEAFOKLNPCPAGIIDLQSVIPNSNE 405
Db 291 SVNWPEVPLDNIISILQAKMPRPLKPHVCTSLVSAWSIMDPPPVGILQOLSKRITES-- 348
Qy 406 PLKVTASVTIPAGRLVCGREGDSIVIVPATQTAIVQLLQGEHMLRRGPPHRTLGRHN 465
Db 349 PVKLTSSIVLPQOSRLVIGREGDSIVIVPATQVMMQLLVGIKQNFSDWPSHQLYCHRG 408
Qy 466 KVTCLLYPHQVSARYDQRYLISGVDPFSVIIWDIFSGEMKHIPCVNGGEITQLLVPPENC 525
Db 409 RVNCLLCPSMIHSRYEKSHLLSGGIDFAVCLMDLYSGSLHHRFCVHAGEITQLLVPPESC 468
Qy 526 SARVQHICISVADSHSVGLSLREKCCIMLASHLEPIQVIKWRPDDVVLVWCSDGSVY 585
Db 469 SPRILKICISVADSHSVTLVLSQERKCVTLASRHLFPVVTIKWAPLDDFLVWCSDGSVY 528
Qy 586 VQMDTGALDRCVMGITAIVELINACDE-----AVPAAVDSLSPHVNKLQA 631
Db 529 VQMETGHLDRVLHGLMAEVLVSACDEQAEDEGGSGGSGNGASASEMGWNPVAFVFRG 588
Qy 632 MTRRSALAKNMAHKKLQTLATNLASEASDKGN---LPKYSNLSMVQAIKTNLTDPDI 688
Db 589 LKSRNNNAIR---HATQRTGITOLOQLOQNGNDFLMKHSNPLVIQGLRTPKDAES 644
Qy 689 HVLFPDVEALITQLLTERASRNTALISPENLQKASGSDKGSFLTGKRAAVLFOQVK- 747
Db 645 HILFDDIEGLIFELHSEYYAQMTPATLESGLVHLQPKDGKMHLDASKKIGDFNKKVN 704
Qy 748 -----ETIKENIKHLLDDEDEBEIMR-----QRRESDPEYRSSKSLPTLL 791
Db 705 KAVDVEKILKDXDKGLVQKFKTEIIVEKQVAKVESLQKAVEPHEEQDLASKASKM 764
Qy 792 EYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKLPHCTVSGLLSRGGHSMMLPGY 851
Db 765 EVTHVMEVAQALLSLHSLHNGLOPHLDKMCETRLGLLRPIVPIYSYGVLSKAGYNSLLPTW 824
Qy 852 NQ-----PACKL--SHGKTEVGRKLPASEGV-----GKGTYGVSRAVTTQHLISLSLAN 899
Db 825 QNNYAIPPGIQLPSSSKGKPLPEELQRLHLEHTAVFTSRHLWELSTLTTHNLALVAMSN 884
Qy 900 TLMWMTNATFIGDHMKKGTPRPPRSTPDLKSARGSPSTSSNIVQOQIKQVAAAPVVSARS 959
Db 885 TLLSMAASFL-----PDSEKHK-----KLQRLA-----QRT 911
Qy 960 DADHSGSDPPSAPALHTCFLVNEGNSQLAAMHCWMLPD--LLGLDKFRPPLLEMLARRW 1016
Db 912 DSTLSNEEBREELMAHHISQIKHAWSELLATHCFLLPDKIEALEPKPKRPOVEMVWKGW 971
Qy 1017 QRCLEVREAAQALLLAELERRIEQAGRKEAIDAWAPYLPOYIDHVLISPGVTSEAAQTITT 1076
```

```
Db 972 QHHCIEIREAAQIILLGELTRMGKGRKQLVESMAQVLYPLTYH--TEPIVGAQQQLALIS 1029
Qy 1077 AP-----DASGPEAKV-----QSEHDLVDDDDITTCGL 1104
Db 1030 QPASGGAGSGGNGGVGVGSGGGAGSGGSGVPGGDAHQDEBYEEBEI----- 1084
Qy 1105 SSVPMKKISTSVSEERKQATATVLGVIGABFGAEI--RPPK-----L 1146
Db 1085 -----IRKPSLSSELKRKQTAVILLGVIGABFGQDISPNHRSGISWATGANUTSGV 1139
Qy 1147 LTRPRSSQIPEGFGLTSGGSNSYLARHTCKALTFLLOPPSPKLPHPHSIRRTAIDLIG 1206
Db 1140 AGGERKSSVVEGFGIAN-----NLARLTSMALHLLYAPPSPKLPQYPLREAADLLG 1194
Qy 1207 RGTVTVEPMYDVAVLMLGELLECADEKQOLANTMGLPLSPAADSARSARSHALSATIAR 1266
Db 1195 RGFTVMEPYLDVSKVLLGLEISECG-KAVPNLNYKLPLTPQADACTARHALRLIATAR 1253
Qy 1267 PPAPITTIKAEVHRHTALAAANTOSQONMHT-TTLARAKGEILLRIVIELIKKMPDVTVDLL 1325
Db 1254 PAAPITTMAREVARYTMOQNAGSINTPLTQSVLHKAKGEILQCVEMLDKMQSEIAGLL 1313
Qy 1326 VEVMDIIMYCLEGLVKKGLQBCFPALCRFYMVSVYERNHRIAAGRHGSHVADIRTG 1385
Db 1314 VEVMDIALHCVDNELKNGRLAEALCPAICKFNQISHCAQTRTAVGANGSLAIYELRQN 1373
Qy 1386 KCTIIGHKGPITAVAFADPGRYLATYNTDISHISFWQMTSLGIGMNLNSAPQURCIK 1445
Db 1374 KCOMIPAHPTITSLAFSPDQKYLVSYSCAENRSLFWQSTGFMFG-LGQ-----SQTCTK 1428
Qy 1446 TYQVPPVQPPASPGSHNALKLARLIWTSNRNVLMAHDGKEHRFMV 1490
Db 1429 GYSTAPIDVS--RLNPMRLAKLVWNNRTVTLMLADGSETRFNV 1471

RESULT 14
Q8C711 PRELIMINARY; PRT; 531 AA.
ID Q8C711 AC Q8C711;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Mus musculus 0 day neonate kidney cDNA, RIKEN full-length enriched
DE library, clone.D630037A10 product:hypothetical G-protein beta WD-40
DE repeats containing protein, full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RA The FANTOM Consortium.
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
SEQUENCE FROM N.A.
```

```
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanuki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu M., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tomaru A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: Contains 3 WD repeats.
CC EMBL: AK052730; BAC35119.1; -.
DR InterPro: IPR001680; WD40.
DR InterPro: IPR011046; WD40_like.
DR Pfam: PF00400; WD40; 3.
DR PRINTS: PR00320; GPROTEINBRPT.
DR SMART: SM00320; WD40; 4.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_3.
DR PROSITE: PS00882; WD_REPEATS_2; 2.
DR PROSITE: PS0294; WD_REPEATS_REGION; 2.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 531 AA; 58749 MW; 8D9C5A61E66F0070 CRC64;
```

```
Query Match 34.3%; Score 2667; DB 2; Length 531;
Best Local Similarity 93.9%; Pred. No. 1.1e-169;
Matches 494; Conservative 19; Mismatches 13; Indels 0; Gaps 0;
Qy 1 MAGNSLVLPVLVWGRKAPTHCISAVLLTDDGATIVTGCCHDQICLWDLSELQINPRALL 60
Db 1 MAGNSLVLPVLVWGRKAPTHCISAVLLTDDGATIVTGCCHDQICLWDLSELVNPRL 60
Qy 61 FGHASITCLSKACASSDKQYIVVSASESGEMCLWVSDGRCIEFTYKLAHTGTGTYQFS 120
Db 61 FGHASITCLSKACASGDKRYTVSASANGEMCLWVSDGRCIEFTYKLAHTGTGTYQFS 120
Qy 121 VGNQREGRLCHGHYPEILLVDATSLVLYSVSKISPDWISSMSTIRSHRTQEDTVVAL 180
Db 121 VGNQREGRLCHGHYPEILLVDATSLVLYSVSKISPDWISSMSTIRSHRTQEDTVVAL 180
Qy 181 SVTGILKWIVTSEISDMQDTEPIFEESKPIYCNQCQISFCAPTORSLVVCYSKVRV 240
Db 181 SVTGILKWIVTSEISDMQDTEPIFEESKPIYCNQCQISFCAPTORSLVVCYSKVRV 240
Qy 241 FDAGDYSLLCSPGSENGQWTGDFVSSDKVIWTEGOSYIYKLPASCLPASDSFRSDV 300
Db 241 FDAGDYSLLCSPGSENGQWTGDFVSSDKVIWTEGOSYIYKLPASCLPASDSFRSDV 300
```

QY 301 GKAVENLIPVQHILLDRKDKELLICPPVTRFFVGCYCEYFKLLIQDSSGRINMISD 360
DB 301 GKAVENLIPVQHILLDRKDKELLICPPVTRFFVGCYCEYFKLLIQDSSGRINMISD 360
QY 361 TADKQSEEGGLAMTTSISLOAEFDKLNPCPAGIIDQLSVIPNSNEPLKVTASVYIPAHGR 420
DB 361 IAEKQEADEGLKMTTCISLOAEFDKLNPCPAGIIDQLSVIPNSNEPLKVTASVYIPAHGR 420
QY 421 LVCGREDGSIIVPATQTAIVOLLQGHMLRRGWPPHRTLRGHRNKVTCCLLYPHQVSARY 480
DB 421 LVCGREDGSIIVPATQTAIVOLLQGHMLRRGWPPHRTLRGHRNKVTCCLLYPHQVSARY 480
QY 481 DQRYLISGGVDFSVIWDIFSGEMKHLFCVHGGEITOLLVPPENC 526
DB 481 DQRYLISGGVDFSVIWDIFSGEMKHLFCVHGGEITOLLVPPENC 526

RESULT 15
Q9BL05
ID Q9BL05 PRELIMINARY; PRT; 1335 AA.
AC Q9BL05;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Rab connectin related protein 2.
GN NameRbc-2; ORFNames=Y54F10AM.10;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Ryan E., Courtney L., Yoakum M.;
RT "The sequence of C. elegans cosmid Y54F10AM.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.H.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;

RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RN Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RN Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RG WormBase Consortium;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 6 WD repeats.
DR EMBL; AC025723; AAK29933.2; -.
DR WormBase; WEGene00004314; rbc-2.
DR WormPep; Y54F10AM.10; CE35950.
DR InterPro; IPR001680; WD40.
DR InterPro; IPR011046; WD40_like.
DR Pfam; PF00400; WD40; 6.
DR PRINTS; PR00320; GPROTEINBRPT.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS50082; WD_REPEATS_2; 3.
DR PROSITE; PS50294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 1335 AA; 147103 MW; 19A5418BB1B65E07 CRC64;
Query Match 25.8%; Score 2006.5; DB 2; Length 1335;
Best Local Similarity 31.7%; Pred. No. 8e-125;
Matches 494; Conservative 264; Mismatches 491; Indels 307; Gaps 39;
QY 5 SLVPIVLWGRKAPTHCISAVLLTDDGATVTCCHDQICLWLSVLEQLNPRALLFGHT 64
DB 13 ALSVPVWIGPPEPKNICAIRVLPDGAIIITGAENGHIINWKLAEGLM--PKQLMIGH 70
QY 65 ASITCLSKACASDQKQYIVSASGEMCLMDVSDGRCIEFTKLACTHTGTQYQF-SVGN 123
DB 71 QKITAISQITNTPTNTRFVSASADGRVCLWEIQDGRCIDSTSSIAIHRVIYPTKSRH 130
QY 124 QREGRLCHGHYPEILVVDATSLVLYSLVSKISPDWISMSIIRSHRQEDTVVALSVT 183
DB 131 TRATRLFCIGDYSDIQVMDPQDLTVVFSISRVRPDMWSCFTII-SHPEKEDQLIGMTLS 189
QY 184 GILKWIWVTSSEISDMQDTEPIPEESKPIYCONCOSIFCAFTQRSILVVCVKYRVFDA 243
DB 190 GMMKWMLT-ELEKOPATSLYEDESKRLIEIQRISVSYICSTRRMILLIATCWMIDM 248
QY 244 GDYSLLCGPFSE-NGQTWTGGDFVSSDKVIIWTENGQSYIYKLPASCLPASDSFRSDVGK 302
DB 249 DDLSTVVFHKNHETTKRCVAGYLTDLKVTIGYDSEIHVPQLPIESLQ-----KEGVPP 303
QY 303 AVENLIPPVQHILLDRKDKELLICPPVTRFFVGCYCEYFKLLIQDSSGRINMISD 362
DB 304 VVANNRPTSNY--FDQPN-----PPIVATVNGVQHSRSL--NDVQFAFLPYNSNSL 353
QY 363 DKQGSSEGLAMTTSISLOAEFDKLNPCPAGIIDQLSVIPNSNEPLKVTASVYIP 416
DB 354 ESSSS-----SYEK-----RRFSVVRSSRRDGGVLSNDAVCCSLFVT 390
QY 417 AHGRLVCGREDGSIIVPATQTAIVOLLQGHMLRRGW-----PPHRTLGRHNRKVT 469
DB 391 SQKGLFLGRSDGIVWYACET-----LAQWLRVPAERATSRILTGHDAVRS 439
QY 470 LLYPHQVSARYDORYLISGGVDFSVIWDIFSGEMKHLFCVHGGEITOLLVPPENC 529
DB 440 MFYPFEHDTYDPQYFVSGDDPFSVIVMINSCTRIHRTFTVHGPGVKFSLPSPNSCKQV 499
QY 530 QHCICSVASDHSVGLSLSLREKCKIMLASRHLPIQVTKWRPSDDYLVGCSGSGSVVWQM 589
DB 500 TKCIASLAADNTIALNLIRDSKMLASRHPPIIQVWRPLDDFVLMVLKLDGSGSVVWQM 559

QY 590 DTGALDRCYMGITAVEILNACDE--AVPAAVDSLS-HPAVNLQAMTRRSALAKN-----642
Db 560 ETANLDRIATGLLAEDIMTACDEQI GVEBGTDETSAHHAVQLIRALKNNKMEAVKOKVVG 619
QY 643 -----MAHKLOT-LATNLLASEADKGNLPHYSHNSLMVQAIKTNLTDPDIHVLV 692
Db 620 GSVSGAATPSSHHELTPNCCTAVQLGSPMSILPLFGCAQGAHLVQ-----664
QY 693 FVVEALIIQLLTEASRPNTALISPE---NLQK-----ASGSSDKGSGFLTGKRAAVLFQ 744
Db 665 FEISALIAGILHLDSSAE-----APEGKTTAQKLDAAAAGASDSA-----TGLSRKLTWQ 715
QY 745 QVKETIKENIKHELLDDEEDEIMQRREESPEYRSSKSKPLTLLLEVNLTMDTAKLFM 804
Db 716 -----FEANLYLDVARLML 729
QY 805 SCLHAWGLNEVLDEVCLDRGLMKPHCTVSFGLLSRGHMSLMLP-----GYNQPACKLSH 860
Db 730 SMLHAWCLDEMDVCDKRLSLHRPHQVYFGNVSRQGELSVSLPTRFAADPESFCK---786
QY 861 GKTEVGRKLPASEGVKGKGYGVSRAVTTQHLLSIIISLANTLMSMTNATFIGDHMKKGPR 920
Db 787 -----KSRWQASHSLNTSHLLAVIATNTLMAKNKNSAL-----819
QY 921 PPRPSTPDLKARGSPPTSSNIVQGIKQVAAPVVSARSADADHSGSDPPSAPALHTCFLV 980
Db 820 -----QLAKIRKSVVEQNOSTHQ-----PDRQO-----I 844
QY 981 NEGWSOLAAMHCVMPLD-LIGLDKFRPPLLEMLAREWQRCLEVRERAAQALLIAELRIE 1039
Db 845 KQWSLVAALHCVLLPDHVRPRSSYAPPSIEMLVRRWQDSCLEIREAAQALLIRELTRLG 904
QY 1040 QAGR-----KEAIDAWPYLPQYIDHVTSPGVTSAAQTITTPADASGPEAK 1086
Db 905 AEGRFPPSNDFELKTVRRILIESWTFPLPLDLSLIFGSKLQSSVPTVQPSAPAP---961
QY 1087 VQEEHDLVDDDDITTCGLSSVQMKKIST-----YEERKQATAVLGVIG 1134
Db 962 -----PIPRTKNAPPDVEPVRGSEPTAEGGAGIOQVRRNQATSIILIGVIG 1010
QY 1135 AEFGBAIEPPKLLTRPRSSQIPEGFGLTSGGSNYSLARHTCKALTFLLLQPPSPKLP 1194
Db 1011 SQFGDEL-----NEADLTRATAVSLLELLVAAPSNLIPVH 1045
QY 1195 STIRRTAIDLIGRFTVWPBYMDVSAVLMGLLELCADAOK--OLANITWGLPLSPAADSA 1252
Db 1046 SPLRRAAIDLGRGFVHWEPHLEISKVLGLLDLASNNEKPGTQRRIT-GAPLNAIEDAA 1104
QY 1253 RSARHALSLIATARPAPFITTIKAEVHRHTALANTOSQONMHTTTLARAKGEILRVIEI 1312
Db 1105 RTSQALSLIALARPALITSLSMEVARNA-AAQHOTTOHTVTVSPLLSRSEVLRIIE 1163
QY 1313 LIEKMPDVTVDLLVEYMDIIMYCLEGLSVKKGLOECFFPAICRFYMVSYERNHRIAUGA 1372
Db 1164 LCEKRYNDIIMALLPVGDLIVHCLDITILKHKSSEVFPPIVKFNVAVCSIKRVAFG 1223
QY 1373 RHGSVALYDRTGKQOTIGHKGPITAVAPADPGRYLATYNTDISHISFQWNTSLLSGI 1432
Db 1224 KNGTCVVHELRAKTHSLPSHNGPIAAVAFSEDKYLATYGAEDKGINFPQTSQSLGM- 1282
QY 1433 GMLNSAPOLRECIKTYQVPPQAPASPGSHNALKLARLIWTSNRNVILMAHDGKEHRP 1488
Db 1283 ----GQAQLKLTQSPAPTVPSTPSTGTSFR-PRLVWINSKSLTLMLEPEGREQR 1333

Search completed: June 30, 2005, 17:23:27

Job time : 223 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 17:16:31 ; Search time 195 Seconds
(without alignments)
2947.872 Million cell updates/sec

Title: US-10-645-335-2

Perfect score: 7773

Sequence: 1 MAGNSLVLPVLMGRKAPTH.....TSNRNVLMHDKGHRFMV 1490

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1722976 seqs, 385795295 residues

Total number of hits satisfying chosen parameters: 1722976

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7773	100.0	1490	17	US-10-645-335-2
2	7763	99.9	1490	15	US-10-093-463-154
3	2805.5	36.1	559	15	US-10-094-749-2356
4	1442	18.6	775	15	US-10-108-260A-3494
5	412.5	5.3	1391	16	US-10-437-963-162857
6	228.5	2.9	579	15	US-10-369-493-18893
7	228.5	2.9	608	15	US-10-369-493-20224
8	228	2.9	703	15	US-10-425-114-58817
9	228	2.9	710	16	US-10-425-115-360670
10	220.5	2.8	1005	15	US-10-369-493-20242
11	205.5	2.6	626	15	US-10-369-493-20247
					Sequence 2, Appli
					Sequence 154, App
					Sequence 2956, Ap
					Sequence 3494, Ap
					Sequence 162857,
					Sequence 18893, A
					Sequence 20224, A
					Sequence 58817, A
					Sequence 360670,
					Sequence 20242, A
					Sequence 20247, A

12	204	2.6	891	16	US-10-437-963-153705	Sequence 153705,
13	202	2.6	1136	15	US-10-369-493-19046	Sequence 19046, A
14	197.5	2.5	429	15	US-10-369-493-11348	Sequence 11348, A
15	191.5	2.5	478	15	US-10-369-493-18905	Sequence 18905, A
16	189	2.4	494	16	US-10-478-197-35	Sequence 35, Appl
17	186	2.4	573	9	US-09-925-299-1015	Sequence 1015, Ap
18	186	2.4	573	10	US-09-925-299-1015	Sequence 1015, Ap
19	186	2.4	1356	13	US-10-077-111-10	Sequence 10, Appl
20	185.5	2.4	679	15	US-10-369-493-20059	Sequence 20059, A
21	184.5	2.4	514	11	US-09-987-701-2	Sequence 2, Appli
22	184.5	2.4	514	11	US-09-987-701-4	Sequence 4, Appli
23	184	2.4	721	17	US-10-732-923-7436	Sequence 7436, Ap
24	181.5	2.3	514	11	US-09-987-701-12	Sequence 12, Appl
25	178	2.3	610	15	US-10-369-493-20185	Sequence 20185, A
26	175	2.3	1073	17	US-10-732-923-7538	Sequence 7538, Ap
27	174.5	2.2	465	16	US-10-687-732-10	Sequence 10, Appl
28	174.5	2.2	579	15	US-10-369-493-6573	Sequence 6573, Ap
29	174.5	2.2	587	16	US-10-687-732-25	Sequence 25, Appl
30	174	2.2	478	16	US-10-425-115-240549	Sequence 240549,
31	171	2.2	395	16	US-10-363-829-446	Sequence 446, App
32	171	2.2	981	14	US-10-032-585-7401	Sequence 7401, Ap
33	169.5	2.2	358	17	US-10-732-923-7507	Sequence 7507, Ap
34	169.5	2.2	1140	15	US-10-369-493-19053	Sequence 19053, A
35	169.5	2.2	1246	16	US-10-437-963-169035	Sequence 169035,
36	167.5	2.2	485	14	US-10-132-744A-6	Sequence 6, Appli
37	165.5	2.1	1508	16	US-10-483-506-9	Sequence 9, Appli
38	164.5	2.1	419	15	US-10-425-114-61169	Sequence 61169, A
39	163	2.1	411	16	US-10-437-963-105334	Sequence 105334,
40	163	2.1	1194	9	US-09-876-667-2	Sequence 2, Appli
41	163	2.1	1194	14	US-10-141-618-10	Sequence 10, Appl
42	163	2.1	1205	9	US-09-876-667-16	Sequence 16, Appl
43	162.5	2.1	475	16	US-10-425-115-362607	Sequence 362607,
44	162.5	2.1	557	15	US-10-425-114-72721	Sequence 72721, A
45	162.5	2.1	958	14	US-10-128-714-8398	Sequence 8398, Ap

ALIGNMENTS

RESULT 1
US-10-645-335-2
; Sequence 2, Application US/10645335
; Publication NO. US20050064430A1
; GENERAL INFORMATION:
; APPLICANT: Masakazu Takeuchi
; APPLICANT: Yoshimi Takai
; TITLE OF INVENTION: Rabconnectin-3-binding protein
; FILE REFERENCE: 2003946-0055
; CURRENT APPLICATION NUMBER: US/10/645,335
; CURRENT FILING DATE: 2003-08-21
; PRIOR APPLICATION NUMBER: JP 2002-319521
; PRIOR FILING DATE: 2002-11-01
; NUMBER OF SEQ ID NOS: 4
; SEQ ID NO 2
; TYPE: PRT
; LENGTH: 1490
; ORGANISM: Homo sapiens
US-10-645-335-2

Query Match 100.0%; Score 7773; DB 17; Length 1490;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1490; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAGNSLVLPVLMGRKAPTHCISAVLLTDDGATVGHCHDGOICLWDLVSVEIQINPRALL 60
Db 1 MAGNSLVLPVLMGRKAPTHCISAVLLTDDGATVGHCHDGOICLWDLVSVEIQINPRALL 60
QY 61 FGHASITCLSKACASSDKQIVSASEGEMCLWVSDGRCIEFTKLTACTHTGICQFYQFS 120
Db 61 FGHASITCLSKACASSDKQIVSASEGEMCLWVSDGRCIEFTKLTACTHTGICQFYQFS 120
QY 121 VGNQREGRLCHGHYPILVVDATSLVLYSVLSKISPDWISSMIIIRSHRTQEDTVAL 180
Db 121 VGNQREGRLCHGHYPILVVDATSLVLYSVLSKISPDWISSMIIIRSHRTQEDTVAL 180

Db 121 VGNQREGLLCHGHYPEILLVVDATSLVLSVSKISPDWISSMSIIRSHRTQEDTVVAL 180
Qy 181 SVTGILLKWIIVTSEISDMQDTEPIFEESKPIYVQNCQSISFCAPTORSLVLLVCSKYWRV 240
Db 181 SVTGILLKWIIVTSEISDMQDTEPIFEESKPIYVQNCQSISFCAPTORSLVLLVCSKYWRV 240
Qy 241 FDAGDYSLLCSGSPSENGQWTGDFVSSDKVLIWTENGOSYIYKLPASCLPASDSFRSDV 300
Db 241 FDAGDYSLLCSGSPSENGQWTGDFVSSDKVLIWTENGOSYIYKLPASCLPASDSFRSDV 300
Qy 301 GKAVENLIPVQHILLDRKDKELLICPPVTRFFYGCREFYFKLLIQDSSSGLNIWNISD 360
Db 301 GKAVENLIPVQHILLDRKDKELLICPPVTRFFYGCREFYFKLLIQDSSSGLNIWNISD 360
Qy 361 TADKQSEGLAMTTSISLOEAFDKLNPCPAGIIDQLSVIPNSNEPLKVTASVYIIPAHGR 420
Db 361 TADKQSEGLAMTTSISLOEAFDKLNPCPAGIIDQLSVIPNSNEPLKVTASVYIIPAHGR 420
Qy 421 LVCGREDGSIIVIPATCTAIVOLLQGEHMLRRCWPPHRTLGRHNRKVTCLLYPHOVSARY 480
Db 421 LVCGREDGSIIVIPATCTAIVOLLQGEHMLRRCWPPHRTLGRHNRKVTCLLYPHOVSARY 480
Qy 481 DQYLLISGGVDFSVIIVDFSGEMKHI FVHGGEITQLLVPPENC SARVQHICISVASDH 540
Db 481 DQYLLISGGVDFSVIIVDFSGEMKHI FVHGGEITQLLVPPENC SARVQHICISVASDH 540
Qy 541 SVGLLSUREKCCIMLASRHLFPQIVIKWRPSDDYLVVGCSDGVSYYVQMDTGALDRCVMG 600
Db 541 SVGLLSUREKCCIMLASRHLFPQIVIKWRPSDDYLVVGCSDGVSYYVQMDTGALDRCVMG 600
Qy 601 ITAVEILNACDEAVPAVDSLSHPAVNLKQAMTRRSIALKNMAHKLQTLATNLNLAESA 660
Db 601 ITAVEILNACDEAVPAVDSLSHPAVNLKQAMTRRSIALKNMAHKLQTLATNLNLAESA 660
Qy 661 SDKGNLPKYSHNSLWQAIKTNLTDPDIHVLFPDVEALIIQLLTBESAPNTALISPENL 720
Db 661 SDKGNLPKYSHNSLWQAIKTNLTDPDIHVLFPDVEALIIQLLTBESAPNTALISPENL 720
Qy 721 QKASGSDKGGSPFTGKRAAVLFOQVKETIKENIKHLLDDEBDEEIMRQREESDPEY 780
Db 721 QKASGSDKGGSPFTGKRAAVLFOQVKETIKENIKHLLDDEBDEEIMRQREESDPEY 780
Qy 781 RSSKSPKLTILEYNTWDTAKLFMSCLHAWGLNEVLDCDLRGLMKPCHTVSFGLLSR 840
Db 781 RSSKSPKLTILEYNTWDTAKLFMSCLHAWGLNEVLDCDLRGLMKPCHTVSFGLLSR 840
Qy 841 GGHMSLMLPGYNQPACKLSHGKTEVGRKLPASBVGKGTGVGSRAVTTQHLLSIISLANT 900
Db 841 GGHMSLMLPGYNQPACKLSHGKTEVGRKLPASBVGKGTGVGSRAVTTQHLLSIISLANT 900
Qy 901 LMSMTNATFIGDHMKKGPTPRPSTPDLISKARGSPPTSSNI VQGQIKQVAAPVVSARS 960
Db 901 LMSMTNATFIGDHMKKGPTPRPSTPDLISKARGSPPTSSNI VQGQIKQVAAPVVSARS 960
Qy 961 ADHSGSDPPSAPALHTCFLVNEGWSOLAAHCVMLPDLGLDKFRPPLLEMLARRWQDRC 1020
Db 961 ADHSGSDPPSAPALHTCFLVNEGWSOLAAHCVMLPDLGLDKFRPPLLEMLARRWQDRC 1020
Qy 1021 LEVREAAQALLAEARRIEQAGRKEAIDAWAPYLPQVIDHVISPGVTSEAAQTITAPDA 1080
Db 1021 LEVREAAQALLAEARRIEQAGRKEAIDAWAPYLPQVIDHVISPGVTSEAAQTITAPDA 1080
Qy 1081 SGPEAKVQEEHDLVDDDIITGCLSSVPQMKI STSYEERKQATAILLVGTVGAEFGAE 1140
Db 1081 SGPEAKVQEEHDLVDDDIITGCLSSVPQMKI STSYEERKQATAILLVGTVGAEFGAE 1140
Qy 1141 IEPKLLTRPRSSQIPEGFLTSGGSNYSLARHTCKALTFLLLQPPSPKLPHSTIRRT 1200
Db 1141 IEPKLLTRPRSSQIPEGFLTSGGSNYSLARHTCKALTFLLLQPPSPKLPHSTIRRT 1200
Qy 1201 AIDLIGRGFTWBPYMDVSAVLMGLLELCADAEKQLANITMGLPLSPAADSARSARHALS 1260
Db 1201 AIDLIGRGFTWBPYMDVSAVLMGLLELCADAEKQLANITMGLPLSPAADSARSARHALS 1260

Qy 1261 LIATARPPAFITTTIAKEVHRHTALAANTOSQONMHTTTTLARAKGEILRVIEILIEKMPD 1320
Db 1261 LIATARPPAFITTTIAKEVHRHTALAANTOSQONMHTTTTLARAKGEILRVIEILIEKMPD 1320
Qy 1321 VVDLLVEVMDIIMYCLEGSLVKKKGLQECFPAICRFYVYVYERNHRIAVGARHGSVALY 1380
Db 1321 VVDLLVEVMDIIMYCLEGSLVKKKGLQECFPAICRFYVYVYERNHRIAVGARHGSVALY 1380
Qy 1381 DIRTGKQTIHGHKGPIITAVAPADGRYLATVNTDSDHISFWQMTSLLSIGMLNSAQ 1440
Db 1381 DIRTGKQTIHGHKGPIITAVAPADGRYLATVNTDSDHISFWQMTSLLSIGMLNSAQ 1440
Qy 1441 LRCIKTYQVPPVQSPASGSHNALKARLIWTSNRNVILMAHDGKGRFMV 1490
Db 1441 LRCIKTYQVPPVQSPASGSHNALKARLIWTSNRNVILMAHDGKGRFMV 1490

RESULT 2

US-10-093-463-154

; Sequence 154, Application US/10093463

; Publication No. US20030208039A1

; GENERAL INFORMATION:

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Shenoy, Suresh

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Gusev, Vladimir

; APPLICANT: Pochart, Pascal

; APPLICANT: Zhong, Mei

; APPLICANT: Rastelli, Luca

; APPLICANT: Mezes, Peter

; APPLICANT: Smithson, Glennda

; APPLICANT: Guo, Xiaojia

; APPLICANT: Gerlach, Valerie

; APPLICANT: Casman, Stacie

; APPLICANT: Boldog, Ferenc

; APPLICANT: Li, Li

; APPLICANT: Zerhusen, Bryan

; APPLICANT: Tchernev, Velizar

; APPLICANT: Gangolli, Esha

; APPLICANT: Vernet, Corine

; APPLICANT: Pena, Carol

; APPLICANT: Burgess, Catherine

; APPLICANT: Liu, Xiaohong

; APPLICANT: Spytek, Kimberly

; APPLICANT: Gorman, Linda

; APPLICANT: Spaderma, Steven

; APPLICANT: Voss, Edward

; APPLICANT: Malyankar, Uriel

; APPLICANT: Anderson, David

; APPLICANT: Patturajan, Meera

; APPLICANT: Miller, Charles

; APPLICANT: Taupier, Raymond J. Jr.

; TITLE OF INVENTION: No. US20030208039A1e1 Antibodies that Bind to Antigenic Polypepti:

; FILE OF INVENTION: Encoding The Antigens, and Methods of Use.

; CURRENT APPLICATION NUMBER: US/10/093,463

; CURRENT FILING DATE: 2002-06-24

; PRIOR APPLICATION NUMBER: 60/283,675

; PRIOR FILING DATE: 2001-04-14

; PRIOR APPLICATION NUMBER: 60/338,092

; PRIOR FILING DATE: 2001-12-03

; PRIOR APPLICATION NUMBER: 60/274,281

; PRIOR FILING DATE: 2001-03-08

; PRIOR APPLICATION NUMBER: 60/274,101

; PRIOR FILING DATE: 2001-03-08

; PRIOR APPLICATION NUMBER: 60/325,681

; PRIOR FILING DATE: 2001-09-27

; PRIOR APPLICATION NUMBER: 60/304,354

; PRIOR FILING DATE: 2001-07-10

; PRIOR APPLICATION NUMBER: 60/279,995

; PRIOR FILING DATE: 2001-03-30

; PRIOR APPLICATION NUMBER: 60/294,899

; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 154
; LENGTH: 1490
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-093-463-154

Query Match 99.98; Score 7763; DB 15; Length 1490;
Best Local Similarity 99.99; Pred. No. 0;
Matches 1488; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	1	MAGNSLVLPVLMGRKAPTHCISAVLITDDGATIVTGDHGOICLWDLSELQINPRALL	60
DB	1	MAGNSLVLPVLMGRKAPTHCISAVLITDDGATIVTGDHGOICLWDLSELQINPRALL	60
QY	61	FHTASITCLSKACASSDKQYIVSASESGEMCLWDVSDGRICBFTKLACTHTGIGYQFS	120
DB	61	FHTASITCLSKACASSDKQYIVSASESGEMCLWDVSDGRICBFTKLACTHTGIGYQFS	120
QY	121	VGNOREGLLCHGYHPEILVVDATSLVLYSLYSKISPDWISSMSIIRSHRTOEDTVVAL	180
DB	121	VGNOREGLLCHGYHPEILVVDATSLVLYSLYSKISPDWISSMSIIRSHRTOEDTVVAL	180
QY	181	SVTGILKWTIVTSEISDMQDTEPIFBEESKPIYVCONQOSISFCAPTORSLLVVCSKYRW	240
DB	181	SVTGILKWTIVTSEISDMQDTEPIFBEESKPIYVCONQOSISFCAPTORSLLVVCSKYRW	240
QY	241	FDAGDYSLLCSGSPENGQWTGDFVSSDKVLIITWENGQSYIYKLPASCLPASDSFRSDV	300
DB	241	FDAGDYSLLCSGSPENGQWTGDFVSSDKVLIITWENGQSYIYKLPASCLPASDSFRSDV	300
QY	301	GKAVENLIPVQHILLDRKDKELLICPPVTRFFYGCREFPHKLLIQDSSGRINIWNISD	360
DB	301	GKAVENLIPVQHILLDRKDKELLICPPVTRFFYGCREFPHKLLIQDSSGRINIWNISD	360
QY	361	TADQSGSEGLAMTTSISLQEAFLKNCPCAGIIDLVSIPNSNEPLKVTVASVYIPAHR	420
DB	361	TADQSGSEGLAMTTSISLQEAFLKNCPCAGIIDLVSIPNSNEPLKVTVASVYIPAHR	420
QY	421	LVCGRDGSIVIVPATQTAIVOLLQGEHMLRRGPPHRTLGRHNKVTCLLYPHQVSARY	480
DB	421	LVCGRDGSIVIVPATQTAIVOLLQGEHMLRRGPPHRTLGRHNKVTCLLYPHQVSARY	480
QY	481	DQYILSGGVDFSVIIVDIPSGEMKHIFCVHGGEITQLLVPPNCARSARVHCICSVASDH	540
DB	481	DQYILSGGVDFSVIIVDIPSGEMKHIFCVHGGEITQLLVPPNCARSARVHCICSVASDH	540
QY	541	SVGLLSIREKKCTMLASRHLFPQIVLKWPSDDYLVVCGSDGSVYVQMDTGALDRCVMG	600
DB	541	SVGLLSUREKKCTMLASRHLFPQIVLKWPSDDYLVVCGSDGSVYVQMDTGALDRCVMG	600

QY	601	ITAVEILNACDEAVPAVDLSLHPAVNLKQAMTRSLAALKNMAHKLQTLATNLASEA	660
DB	601	ITAVEILNACDEAVPAVDLSLHPAVNLKQAMTRSLAALKNMAHKLQTLATNLASEA	660
QY	661	SDKGNLPKYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIQLITEASRPNTALISPENL	720
DB	661	SDKGNLPKYSHNSLMVQAIKTNLTDPDIHVLFPDVEALIIQLITEASRPNTALISPENL	720
QY	721	QKASGSSDKGSGFLTCGRAAVLFQVKETIKENIKHLLDDEEDEEIMRQRREESDPEY	780
DB	721	QKASGSSDKGSGFLTCGRAAVLFQVKETIKENIKHLLDDEEDEEIMRQRREESDPEY	780
QY	781	RSSKSKPLTILEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHCTVSGLLSR	840
DB	781	RSSKSKPLTILEYNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKPHCTVSGLLSR	840
QY	841	GGHMSLMLPGYNQPACKLSHGKTEVGRKLPASEGVGKGTGVSVRAVTTQHLLSIIISANT	900
DB	841	GGHMSLMLPGYNQPACKLSHGKTEVGRKLPASEGVGKGTGVSVRAVTTQHLLSIIISANT	900
QY	901	LMSMTNATFIGDHMKKGTPRPPRPTDLSKARGSPPTSNNIVQGGIKQVAAAPVVSARS	960
DB	901	LMSMTNATFIGDHMKKGTPRPPRPTDLSKARGSPPTSNNIVQGGIKQVAAAPVVSARS	960
QY	961	ADHSGSDPPSAPALHTCFLVNEGWSQLAAHVCMLPDLGLDKFRPILMLARRQDRC	1020
DB	961	ADHSGSDPPSAPALHTCFLVNEGWSQLAAHVCMLPDLGLDKFRPILMLARRQDRC	1020
QY	1021	LEVREAAQALLAELRLRIEQAQRKEAIDAWAPVLPQYIDHVISPGVTSAAQTITAPDA	1080
DB	1021	LEVREAAQALLAELRLRIEQAQRKEAIDAWAPVLPQYIDHVISPGVTSAAQTITAPDA	1080
QY	1081	SGPEAKVQSEEDHLDVDDDDITTCGLSSVPQMKKISTSYEERRKQATAIIVLGVIGAFGAE	1140
DB	1081	SGPEAKVQSEEDHLDVDDDDITTCGLSSVPQMKKISTSYEERRKQATAIIVLGVIGAFGAE	1140
QY	1141	IBPPKLLTPRSSSQIPEGFGLTSGGSNYSLARHTCKALTFLLLQPPSPKLPHPSTIRRT	1200
DB	1141	IBPPKLLTPRSSSQIPEGFGLTSGGSNYSLARHTCKALTFLLLQPPSPKLPHPSTIRRT	1200
QY	1201	AIDLIGRGFTVPEYMDVSAVLMGELLECADAEKQALANTMGLPLSPAADSARSARHALS	1260
DB	1201	AIDLIGRGFTVPEYMDVSAVLMGELLECADAEKQALANTMGLPLSPAADSARSARHALS	1260
QY	1261	LIATARPAPAFITTIKAEVHRHTALAAANTOSQOQNMHTTTLARAKGEILRVIELIEKMPD	1320
DB	1261	LIATARPAPAFITTIKAEVHRHTALAAANTOSQOQNMHTTTLARAKGEILRVIELIEKMPD	1320
QY	1321	VVDLLVEVMDIIMYCLEGLSVKKGLOECFPAICRPFYVSYERNHRIAVGARHGSVALY	1380
DB	1321	VVDLLVEVMDIIMYCLEGLSVKKGLOECFPAICRPFYVSYERNHRIAVGARHGSVALY	1380
QY	1381	DIRTKGCTQIHGHKGPIITAVAPDGRYLATYNTDISHISFWQMTSLGSGIMLNSAQ	1440
DB	1381	DIRTKGCTQIHGHKGPIITAVAPDGRYLATYNTDISHISFWQMTSLGSGIMLNSAQ	1440
QY	1441	LRCIKTYQVPPVQSPAGSHNALKLARLWTSNRNVLMAHDKGKEHRFMV	1490
DB	1441	LRCIKTYQVPPVQSPAGSHNALKLARLWTSNRNVLMAHDKGKEHRFMV	1490

RESULT 3

US-10-094-749-2956
; Sequence 2956, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROFUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI

APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: TRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHICO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749
PRIOR FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2956
LENGTH: 559
TYPE: PRT
ORGANISM: Homo sapiens
US-10-094-749-2956

Query Match 36.1%; Score 2805.5; DB 15; Length 559;
Best Local Similarity 94.3%; Pred. No. 1.1e-229;
Matches 549; Conservative 0; Mismatches 0; Indels 33; Gaps 1;
QY 526 SARVQHCICSVADSHVGLLSREKKCIMLASRHLPIQVWKRPDDVYVCCSGSVY 585
DB 11 SARVQHCICSVADSHVGLLSREKKCIMLASRHLPIQVWKRPDDVYVCCSGSVY 70
QY 586 VQWMDTGALDRVCWGITAVEILNACDEAVPAADVSLSHPAVNLQAWTRRSAAALKNMAH 645
DB 71 VQWMDTGALDRVCWGITAVEILNACDEAVPAADVSLSHPAVNLQAWTRRSAAALKNMAH 130
QY 646 HKLQTLATNLLASEADSGKGNLPKYSNSLMVQAIKTNLTDPDTHVLFFDVEALIIQLLTE 705
DB 131 HKLQTLATNLLASEADSGKGNLPKYSNSLMVQAIKTNLTDPDTHVLFFDVEALIIQLLTE 190
QY 706 EASRPNTALISPENLQKASGSSDKGSLFTGKRAAVLFOQVKETIKENIKEHLDDDEED 765
DB 191 EASRPNTALISPENLQKASGSSDKGSLFTGKRAAVLFOQVKETIKENIKEHLDDDEED 250
QY 766 EETMRORRESDPEYRSSKSLPTLLEYNLMTDTAKLFMSCLHAWGLNEVLDDEVCLDRLG 825
DB 251 EETMRORRESDPEYRSSKSLPTLLEYNLMTDTAKLFMSCLHAWGLNEVLDDEVCLDRLG 310
QY 826 MLKPHCTVSPGLLSRGHMSLMLPGYNQPACKLSHGKTEVGRKLPASEGVGKGTGVGSRA 885
DB 311 MLKPHCTVSPGLLSRGHMSLMLPGYNQPACKLSHGKTEVGRKLPASEGVGKGTGVGSRA 370
QY 886 VTTHLLSIISLANTLMSMTNATFIGDHMKKGTPRPRSTPDLKARGSPPTSSNIVQ 945
DB 371 VTTHLLSIISLANTLMSMTNATFIGDHMKKGTPRPRSTPDLKARGSPPTSSNIVQ 430
QY 946 QIKQVAPVVSARSADSHSGSDPPSAPALHTCFLVNEGWSQLAAMHCWMLPDLGLDKFR 1005
DB 431 QIKQ-----GWSQLAAMHCWMLPDLGLDKFR 457
QY 1006 PPLEMLARWQRCLEVRERAAQALLAEARRIEQAGRKEAIDAWAPYLPQYIDHVISPG 1065
DB 458 PPLEMLARWQRCLEVRERAAQALLAEARRIEQAGRKEAIDAWAPYLPQYIDHVISPG 517
QY 1066 VTSEAAQTITAPDASGPEAKVQEEHDLVDDITTGCLSSV 1107
DB 518 VTSEAAQTITAPDASGPEAKVQEEHDLVDDITTGCLSSV 559

US-10-108-260A-3494
Sequence 3494, Application US/10108260A
Publication No. US20040005560A1
GENERAL INFORMATION:
APPLICANT: HELIX RESEARCH INSTITUTE
TITLE OF INVENTION: No. US20040005560A1el full length cDNA
FILE REFERENCE: H1-A0106
CURRENT APPLICATION NUMBER: US/10/108,260A
CURRENT FILING DATE: 2002-03-27
NUMBER OF SEQ ID NOS: 5458
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3494
LENGTH: 775
TYPE: PRT
ORGANISM: Homo sapiens
US-10-108-260A-3494

Query Match 18.6%; Score 1442; DB 15; Length 775;
Best Local Similarity 38.6%; Pred. No. 4.5e-113;
Matches 303; Conservative 173; Mismatches 268; Indels 40; Gaps 11;
QY 10 IVLWGRKAPTHCISAVLLTDDGATVTCGHDQOICLWDLSELQINPRALLFGHTASITC 69
DB 8 VALWGQKAPPHSITAITMITDDQRTTIVTGSQEGQLCWNLSHELKTSAKELLFGHSASVTC 67
QY 70 LSKACASSDKQYIVSAGESGEMCLWDVSDGRICIEFTKLACTHTGTQFYQFSVGNQREGRL 129
DB 68 LARARDFSKQPIYVSAENGEMCVWNVINGQCVKATLPYRHTAICYHCSFRMTGEGWL 127
QY 130 LCHGHPYELIVVDATSLVLYSLVSKISPDWISSMSIIRSHRTQBDTVVALSVTGILKVM 189
DB 128 LCCGEYQDVLIIDAKTLAVVHSFRSSQPPDWINCMCIHVSMRIQEDSLVWSVAGELKVM 187
QY 190 IVTSEISMDQTEPIFEESKPIYCONQOSISFCAPTQBSLLVVCCKWRFVDAGDYSLL 249
DB 188 DLSSINSIQERQDVYEKESFLESINCOITIRFCTYTERLLLVVFSKCMKVVDYCDLSLL 247
QY 250 CSGSPENGQWVGDFVSSDKVIIWTEGOSYIYKLPASCLPASDFSFRSDVCKAVENLTP 309
DB 248 LTEVRNGOFFAGGEVIAHRIIWTEDGHSYIYQLLSGL--SKSIYPADGRVLKETIY 305
QY 310 PVQHIILD---RKOKELLICPPVTRFFYGCREYFHKLLIQGSSGRLNIWNISD--TADK 364
DB 306 P--HLLCSTSVQENKE-QSRPFVGMVNERKEPFYKVLFSGEVSGRIITLWHIPDPVPSKF 362
QY 365 QSGSEGLAMTTSISIQEAPDKNCPAGIIDQLSVIPNSNEPLKVTASVYIIHAHGLVCG 424
DB 363 DGSPREIPVTATWTLQDNFDKHTMSQSIIDYFSGLDKGAGTAVVTSSEYIYPSLDKLCG 422
QY 425 REDGSIIVIPATQTAIVQLLQGEHMLRRCMPPHRTLRGHRNKVTCLLYPHOVSARYDQRY 484
DB 423 CEDGTIIITQALNAKARLEGGSIKVS--PHKVLKHHQSVTSLLYPHGLSSKLDQSW 481
QY 485 LISGVDVFSVIIWIDIFSGEMKHI FCVHGGEITQLLVPPENCARSARVQHCTCSVASHSVGL 544
DB 482 MLSGDLSDCVILWDITFTEILHKFLEAGPVTSLMSPEKFKLRGEQIICCVCGDHSVAL 541
QY 545 LSIREKKCIMLASRHLFPQVTKWRPSDDYLVGSGDSGVYVQWMDTGALDRCVWGITAV 604
DB 542 LHLEGKSCILLHARKHLFPVRMIKHPVENFLVGCADDSVYIWEIETGTLEHRTGERAR 601
QY 605 ETLNACDEA-----VPAADVDSLHPAVNLKQAWTRRSAAALKNMAHKLQTLATNLLAS 658
DB 602 IILNCCDDSQLVKSVLPITASETLKHKSTEQRS-----SPYQLGFLPCPLQV 649
QY 659 EASDKGNLPKYSNSLMVQAIKTNLTDPDTHVLFFDVEALIIQLLTEAS--RPNTALIS 716
DB 650 ESSCKVTDKFCPRPNVLPVKTWNSVGFHILLFDLENLVELLLPTPLSDVDSSSFFYG 709
QY 717 PENLQKASGSSDKGSLFTGKRAA-----VLFOQVKETIKENIKEHLDDDEEDEM 769
DB 710 GEVLRRAKSTVEKKTITLTKSKTACGPLSAAELAKPITESLAQG--DNTIKFSEENDGIX 767

QY 770 RQR 773
Db 768 RQK 771

RESULT 5
US-10-437-963-162857
; Sequence 162857, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 162857
; LENGTH: 1391
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_61908C.1.pap
US-10-437-963-162857

Query Match 5.3%; Score 412.5; DB 16; Length 1391;
Best Local Similarity 19.5%; Pred. No. 8.9e-25;
Matches 340; Conservative 210; Mismatches 560; Indels 631; Gaps 81;

QY 12 LWRKAPTHCISAVLLTDDGATVGTGCHDQICLWLSVBLQINPR--ALLFGHTASIT- 68
Db 9 LWSPPSPSHITAAATP--AALTGAADGTILHWPLPPSPSPSSILLCAHAAIT 66
QY 69 -----CLSKACASSDKQYIVSASEGEMCLWVSDGRCIEFTKL----- 107
Db 67 LCLPSPSPCLLASCAGVLS-LFSSASASASL-----RCLRRSLPWPAGSPSLVAP 119
QY 108 ---ACTHTGTFQYQFVSGNOREGLCHG-----HYPEILVVDATSLV-----YSLVS 154
Db 120 LPSSSSAG-----SSASVAILCHAPDDGGRHVAVVVDARTLVLRAPFGALS 171
QY 155 KISPDWI-----SSMIIIRSHRTQEDTVVAL-----SVTGILKVVITVTEISDMQD 200
Db 172 VAPPRALAVADAGVEDASVSVVLADAGRAQVVPVVAEGAAGVGDSPRRLSASSASVTS 231
QY 201 TEPI-----PEESKPIYQNCQISFCATQSRLLVVCCKYWRVFDAGDYS-----L 248
Db 232 AEAVDGRVAVSLSDGKV-----ALVMKNSCLLKISSEGVL--GEVTLPSDL 279
QY 249 LCGSPSENGQW--TGDFVSSDK-----VIWTENGOSIYIKLPASCLPAS 293
Db 280 LCKEGERGMKGLVGGFLLRGGEWGAHSGENNVRSVLVMSINGAIVY----- 329
QY 294 DSRPSDVG-----KAIVENLIPVQHILLDRDKKELL-ICPPVTRPFYGCYFHKLLIQ 346
Db 330 ---RVEVGTSGFGKAV---CEIPDIVSBERGDSLVQFCQGNQ-----LIR 370
QY 347 GDS-----SRL-----NITSDTK-----QSEBGLAMTTSISLO 380
Db 371 VESRPYKIASLLWKPFSVSIWMDHLELNTANNIEKPLPSKILGEGLOEEFRSDHS-- 428
QY 381 EAPDKLNPCCPAGI-IDOLSVIPNSN-----BPLKVTSVYIPAHGRVLCGREDGS 429
Db 429 HSFQSN---NGVDINSLICSSNSGLGRGGTVSSMWLSSESYTPY--AVVYGFHGD 483
QY 430 IVIVPATQTAIVQLQGEHMLRRGWPPH---RTLGRHNRKVTCLLYPH---QVSARYDOR 483

Db 484 IEVI-----RFLNLLPAKFGSGGIYPHISERFPLGHTGAILCLAAHMHQAOPDSRTFNR 538
QY 484 YLISGGVDFSVIWDIFSGEMKHIFCVHGGEITQLLVPPENCARSARVOHCISCVASDHSVG 543
Db 539 VLISGSFDSITRVWDLDAAGTILSVMHVAVPVQIMLPANWTHQPDWDCFLSVGEIGIVA 598
QY 544 LLSREKKKICIMLASRHLFPQIVIKWRPSDDYLV-----GCSG-SSV-YVWQMDTGALD 595
Db 599 LNSLQTRVVERMPFGHPGSPVSMVMDGVKGYIACLCLRNHLSCNDSGLVYIMDLTKTCARE 658
QY 596 RCVMGITAIVELL-----NACDEAVPAVDLSHPAVNL--KQATRRSLAALKNMA-- 644
Db 659 RIITGTSOSSTFEHFCRGISKNAVTGSIIGTTSASSLLVPPIPKDTSLLQSHANKGLSI 718
QY 645 -----HHKLQTLATNLASEASD--KGNLPKYSH-----NSLMVQAI 679
Db 719 SSVSTVNHANVTNSVTPASADVGMKMSATDEAHELHNSGSKVASGOCINRRKHPI 778
QY 680 KTNLTDPDIVHFFDVEALIIQLLBEASRPNTALISPENLQKAGSSDKGGSFLTGRKA 739
Db 779 KCSCPYPGIASLRFDL-----TAIMSTQGM--ANNNSDR----- 810
QY 740 AVLFOQVKETIKENIKHEILLDBEEDDEEIMRQREES-----DPEYSSSKPKPILILE 792
Db 811 -----QLRDHFYRDVND--SIQAECDNTSGMVIDSPSRESLEGR---LLR 853
QY 793 YNLTMDTAKLFMSCLHAWGLNEVLDEVCLDRGLMQLPH--CTVSFGLSLRGHSMMLPGY 851
Db 854 FSLCP-----LHLGVDELHDLKLVDEMQVCKEPCCHIATGVVGRGSLTFMPPG- 903
QY 852 NOPACKLSHGKTEVGRKLPASEGVGKGYGVSRAVTTQHLLSIIS--LANTLM--SMTNATF 909
Db 904 -----KEATLEIWSESSAKCLMDNQVNAIL 927
QY 910 IGDHMKKGTRPRRSTPD-----LSKARGSPPTSSNIVOGQIKQVAPVVSARSADHSG 965
Db 928 MDLQLLKNKVPFQSLPTSQDMNIIITAIQASVSS---YGOLKADNEDV--GREDCDTSE 982
QY 966 SPPSPAPALHTCFLVNEGWSQLAAHCVMLPDLGLDKP--RPPLLEMLARRWODRCLEVR 1024
Db 983 I-----SSW-----LESFENQEWLSWIGTSQD----- 1005
QY 1025 EAAQALLAELRRIEQAQRKEAIDAWAP-----YLPQVIDHVISPG--VTSEAAQ 1072
Db 1006 AVASNIIVA-----AALVWVPSIVKPKLAHLVVNQIILKLVMSNDRYSSTAEE 1054
QY 1073 TITPADSAPRAKQVEEHLVDDDDITTYGCLSSVPQMKKISTSYEERRKQATAIVLLGV 1132
Db 1055 LLAEGMESTWKVCLGTDMTFLSDVLFOIECLSSAPSNNAV-----YKTAVAVTMR-- 1105
QY 1133 IGAEFGAETEPKLLTRPRSSQIPGFGLTSGGSNSYLARHTCKALTFLLLQPPSPKLP 1192
Db 1106 -----EALVGTLL-----PSL- 1116
QY 1193 PSHITRTAIDLIGRGF-----TWPEYMDVSAVLMGLLELCAEAQKOLANITMGLP-- 1244
Db 1117 -----AMADIV--GFFGVIOQIWAATSSDSPVHVISL-----KTLIRVVRSPKA 1159
QY 1245 LSPAADSARSARHALSLIATARPAPITTIKAEVHRHTALAAANTOSQONNHTTLARAKG 1304
Db 1160 LAPYLDK-----AISYV-----LHT----- 1174
QY 1305 EILRVIEILLERKPTDVVLLVEMDIIM--YCLBGLSVKKKGLQECFPAICRFYVMVSY 1362
Db 1175 -----MDPSNLINRKACIINSMALREIARVFP-----MVALN 1207
QY 1363 ERNHRIAVG-----ARHGSVALYDITRGCTQIHGKGP-----ITA 1399
Db 1208 ESMTRLAVGDALGEIHNAITRVYDIESVTKIRILDASGPPGLPSLLDGSNTTATILITA 1267
QY 1400 VAFAPDGRVLYATYSNTDISHISFWQNTSLGSGIMLSNAPQRCITKYQVPPVQSPASPS 1459
Db 1400 VAFAPDGRVLYATYSNTDISHISFWQNTSLGSGIMLSNAPQRCITKYQVPPVQSPASPS 1459

Db 1268 LSFLEGEGLVAFSENGLMIRWNSLGNWWRLS--RSLTPIQTKLIYVPPWMEGSPNS 1325
QY 1460 -----HNAKLABLIWTSNRNVILMAHDKG 1484
Db 1326 ARLSIISILGHDKHONSETKTRELDEADNLKLLAHNLDSLRYLQWVGSKTIKLTTRH-GQ 1384
QY 1485 E 1485
Db 1385 E 1385

RESULT 6
US-10-369-493-18893
; Sequence 18893, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18893
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Anabaena PCC7120
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(579)
; OTHER INFORMATION: unsure at all xaa locations
US-10-369-493-18893

Query Match 2.9%; Score 228.5; DB 15; Length 579;
Best Local Similarity 20.6%; Pred. No. 9.8e-10;
Matches 120; Conservative 80; Mismatches 203; Indels 179; Gaps 24;
US-10-369-493-18893

QY 19 TGCISAVLLTDDGATIVTGCCHDQICLWDLVSVELQINPRALLFGHTASITCLSKACASSD 78
Db 109 TSCVRSVVFSDGAMLASGSDQTVRLWDIS---SGNCLYTLQHT---SCVRSVVFSPD 162
QY 79 KQIVSASEGEMCLWDVSDGRGIEFTKLACTHTGIGFYQFVSGNOREGRLCHGYPEI 138
Db 163 GAMLASGDDQIVRLWDISSGNCL-YTLOGYT-SWVRFLVFS---PNGVTLANGSSDQI 216
QY 139 L-VVDATSLVLSVLSKISPDWISSMSIIRSHRTQEDTVVALSVTG--ILKVIWTVTSEI 195
Db 217 VRLWDISSKKCLYTLQ--HTNWNVAFA-----SPDGATLASGSDQTVRLWDISS-- 266
QY 196 SDMQDTEPIFEESKPIYQNCOSISFCAPTORSLLVCSKYW---RVPDAGVSYLLCSG 252
Db 267 -----SKCLY-----ILQHTSWNSVVFNP-DGSTLASG 295
QY 253 PSENGQTWGTGDFVSSDKVI-IWTENGQSYIYKLPASCLPASDSFRSDVKAVENLIPPV 311
Db 296 -----SSDQTVRLWEINS-----SKCLCTFQHTSWNSVVFNP----- 328
QY 312 QHILLDRKOKELLICPPVTRFFYGCREFYFKLLIQDSSGRNLNWNISDTADKQGSBEGL 371
Db 329 -----PD-----GSMLASGSSDKTVRLWDISSSKCLLHTFQHT 361
QY 372 AMTTSISLQEAFLKLPACAGIIDQLSVIPNSNEPLKVTASVIPAHGRVLCVREGDSIV 431
Db 362 NWNSVA-----FNP-----DQSM 376
QY 432 IV-PATQTAIVQLLQGEHMLRRGWPPHRTLRGHRNKVTCILLYPHOVSAARYDQRYLISGGV 490

Db 377 ASGSGDQTVRLWEISSKCL-----HTFQHTSWNS-----SVTFSPDGTMLASGSD 423
QY 491 DFSVIWIDFSGEMKHIFCVHGGGEITQLLAVPPENCARSVQHICICSVASDHSVGLLSLREK 550
Db 424 DQTVRLWISSGECUYYTFLGHTNWWGSVIFSPDGA-----ILASGGDQTVRLWISSG 477
QY 551 KCIMLASRHLPIQVIKWRPSDDYLVVCGSDGSVVWQMDTG 592
Db 478 KCLYTLQGHNNWVGSI VSPDGTLLASGSDQTVRLWNISSG 519

RESULT 7
US-10-369-493-20224
; Sequence 20224, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20224
; LENGTH: 608
; TYPE: PRT
; ORGANISM: No. US20030233675A1toc punctiforme
US-10-369-493-20224

Query Match 2.9%; Score 228.5; DB 15; Length 608;
Best Local Similarity 18.4%; Pred. No. 1.1e-09;
Matches 112; Conservative 89; Mismatches 176; Indels 231; Gaps 25;
US-10-369-493-20224

QY 23 SAVLLTDDGATIVTGCCHDQICLWDLVS-ELQINPRALLFGHTASITCLSKACASSDKQY 81
Db 27 AGIAPSPDGTLLATDAEGELRLWEVATGKLVN----FAGHLGWVWSLA---FSPDQGL 79
QY 82 IVSASBSGEMCLWDVSDGRGIEFTKLACTHTGIGFYQFVSGNOREGRLI----- 130
Db 80 LASCSSDKTIRLWDVNTGKCLR-----TLSGHTSSINWSVAFSADQOMLASGDEBTIRL 133
QY 131 -----CH-----GHYPEILVVDATSLVLYSLVSKISPDWISSMSIIRSHRTQEDTVVA 179
Db 134 MNVNTGDCHKIFSGHTDRILSL-----SFSSDGTLLAS 166
QY 180 LSVTGILKVIWTVTSEISMDQDTEPIFEESKPIYQNCOSISFCAPTORSLLVCSKYWR 239
Db 167 GSADFTIRLWKISGE-----CDRILEGHSDRIW-----SISF----- 198
QY 240 VPDAGVSYLLCSGSPENGQTWGTGDFVSSDKVI-IWTENGQSYIYKLPASCLPASDSFRS 298
Db 199 -----SPDQGLVSG---SADFTIRLWEVS-----TGNCF----- 225
QY 299 DVGKAVENLIPPVQHILDRKD--KELLICPPVTRFFYGCREFYFKLLIQDSSGRNLNIV 356
Db 226 -----NILQHSDRVRSALAFSPA-----QMLVSASDDKTVRIW 259
QY 357 NISDTRADKQGSBEGLAMTTSISLQEAFLKLPACAGIIDQLSVIP-NSNEPLKVTASVII 415
Db 260 EAS-----TGEC-----LNLPGHTSIFSVAFNV-- 284
QY 416 PAHGRVLCVREGDSIVIPATQTAIVQLLQGEHMLRRGWPPH-----RTLGRHNRKVTCL 470
Db 285 -----DGRFTASGSDQTV-----KLWDVNTGRCFKTLKGYNSV--- 319
QY 471 LYPHOVSAARYDQRYLISGGVDFSVIWDIFSGEMKHI FCVHGGGEITQLLAVPPENCARSVQ 530

```
Db 320 ---FSVAFNLDGQTASGSDTQTVRLWDVNTGCTKKFAGHSGWTVSVAFHPDG----- 370
Qy 531 HCICSVASDHSVGLSLREKCMILASRHLFPQIVIKWRPDDYLYVVGCSGSGVYVWQMD 590
Db 371 DLLASSADRITRLWSYSTGCLQLKXDHVNWQSVAFSPDRQILASGSDDTQIRLWSVS 430
Qy 591 TGAIDRCV 598
Db 431 TG---KCL 435

RESULT 8
US-10-425-114-58817
; Sequence 58817, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 58817
; LENGTH: 703
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700086365_FLI.pep
US-10-425-114-58817

Query Match 2.9%; Score 228; DB 15; Length 703;
Best Local Similarity 17.9%; Pred. No. 1.5e-09;
Matches 166; Conservative 126; Mismatches 273; Indels 360; Gaps 33;

Qy 671 HNSLMVQAIKTNLTDPDIHVPDVEALIIQLTEASRPNTALISPENLQKASGSDKG 730
Db 16 HNNKIKCPKICSPYPGIASLRFDL-----TAIMSTQGM--TNSNSDR- 56
Qy 731 GSFLTGKRAAVLFOQVKETIKENIKHLLDDEDEEIMR-----QRREESDPEYRSS 783
Db 57 -----HLKDHLL--HSENPKEMLQPGTLDSPSRVHENDSPSRSS 92
Qy 784 KSKPLTLLLYNTMTDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKLPH-CTVSPGLLSRG 842
Db 93 LEG--CLLRFSLCF-----LHLWDVDCLELDKLLVDEMVCQKPEGCHIAATGVVDKG 141
Qy 843 HNSLMVQAIKTNLTDPDIHVPDVEALIIQLTEASRPNTALISPENLQKASGSDKG 902
Db 142 SFTLMPFG-----KEATLELWKS-----SABFCAMRSLSIISLAQLMI 179
Qy 903 SMTNATPIGDHMKKGTPRPRPSTPDLKARGSPPTSSNIVQGIQVAAPVVSARSAD 962
Db 180 KLSRSC-----TNS----- 189
Qy 963 HSGSDPPSAPALHTCFLVNEGWSQLAMH---CVMLPDLLGLDKFRPPLLEMLARWQD 1018
Db 190 -----SALAAYTRHFAEKVPDI-----KPPSLQLLVSWFQW 221
Qy 1019 RCLVREAAQALL-LAELRBIQAGRKEAIDANAPYLQVIDHVPISGVTSE----- 1069
Db 222 PSEHVRMAARSLFHCAAPRSIPQPLRVNK-----NKVNPAPLSSSDMMNVLIC 269
Qy 1070 AATITITPADSGPEAKVQEEHDLVDDITTCGLSSVPQMKKISTSYEERRQKATAVL 1129
Db 270 AVQTASVS-----NHGELKEDDENVDKDDIDT-----ANWILW 302
Qy 1130 LGVIGAEFGAEIBPPKLLTRPRSSQIPEGFLTSGGSNYSLARHTCKALTFFLLQPPSP 1189
```

```
Db 303 L-----ESFENQEWLSWIGTSDAVASNIIVAALVWVPSVV 341
Qy 1190 KLPPHTIRRTAIDLIRGFTVWEPYMDVSAVLMGLLELCADAEKOLANTMG----- 1242
Db 342 KAKLSLVVSQIKLV---MSMDRYSSTAA-----ELLAEGMENTWKAACAEITHFM 392
Qy 1243 -----LPLSPAADSARSARHALSL-----IATARPPAFITTIKAEVHRHT 1282
Db 393 SDILFOIECLSTAPSSNAINKTAIVTMRREALVGTLLPSLAMADVTFGFGVIESQIWA-- 450
Qy 1283 ALAANTOSQONMHTTTLARAKGEILRVEILIEKMTDQVVDLLVVEWMDIMY----- 1334
Db 451 -----TSSDSPVHVASIK-----TIIRV---VRGAPKSLVPYLDKAINYILHTWDPNSLI 497
Qy 1335 -----CLEGSLVKKKGQECFPALCRFVMSVSYERNHRIAVG-----ARHGSVALYDITRTG 1385
Db 498 MRKACIISSMMALREMARVFP-----MWALNESMTKLAVGDAIGEYDGTIRVYDIESV 551
Qy 1386 KCOITHGKGP-----ITAVAPADPGRYLATYNTSDSHISFWMQNTS 1427
Db 552 TKIRILDASGPPGLPSLLAGSSNRTTILISVLSFSPDGBGLVAFSENGLMIRWMSLGSA 611
Qy 1428 LGSIGMNSAPQLRCIKTYQVPPQASPGS----- 1459
Db 612 WWERLS--RSLTPIQCTKLIYVPPWEGFSPNSARLSIISILDKPKRSEKELGEADNLK 669
Qy 1460 ---HNALKLARLIWTSNRNVILMAH 1481
Db 670 LLLHNDLSYRLHWIGGKTIKLRH 694

RESULT 9
US-10-425-115-360670
; Sequence 360670, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 360670
; LENGTH: 710
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_92112C.1.pep
US-10-425-115-360670

Query Match 2.9%; Score 228; DB 16; Length 710;
Best Local Similarity 17.9%; Pred. No. 1.5e-09;
Matches 166; Conservative 126; Mismatches 273; Indels 360; Gaps 33;

Qy 671 HNSLMVQAIKTNLTDPDIHVPDVEALIIQLTEASRPNTALISPENLQKASGSDKG 730
Db 23 HNNKIKCPKICSPYPGIASLRFDL-----TAIMSTQGM--TNSNSDR- 63
Qy 731 GSFLTGKRAAVLFOQVKETIKENIKHLLDDEDEEIMR-----QRREESDPEYRSS 783
Db 64 -----HLKDHLL--HSENPKEMLQPGTLDSPSRVHENDSPSRSS 99
Qy 784 KSKPLTLLLYNTMTDTAKLFMSCLHAWGLNEVLDEVCLDRGLMKLPH-CTVSPGLLSRG 842
Db 100 LEG--CLLRFSLCF-----LHLWDVDCLELDKLLVDEMVCQKPEGCHIAATGVVDKG 148
Qy 843 HNSLMVQAIKTNLTDPDIHVPDVEALIIQLTEASRPNTALISPENLQKASGSDKG 902
```

Db 149 SFTLMFPG-----KEATLELWKS-----SABFCAMRSLSIISLAQLMI 186
Qy 903 SMTNATFIGDHMKGPTRPRPTDLSKARGSPPTSSNIVQOIKQVAPVVSARSAD 962
Db 187 KLSRSC-----TNSS-----196
Qy 963 HSGSDPPSAPALHTCFLVNEGWSQLAAMH-----CVMLPDLGLDKFRPPPLLEMLARRWQD 1018
Db 197 -----SALAAYTRHFAEKVPDI-----KPSLQLLVFSWQH 228
Qy 1019 RCLVREAAQALL-LAELRIEQAQRKEADAWAPYLPQYDHYVSPGVTS-----1069
Db 229 PSEHVRMAARSLFHCAPRSIPQPLRVNK-----NKVVNAPLSSSDNMNVLIC 276
Qy 1070 AAOITITAPDASPEAKVQEEHDLVDYDITTCCLSSVPMKKISISYERRKQATAVL 1129
Db 277 AVQTASVS-----NHGELKEDDENVDKDDIT-----ANMILW 309
Qy 1130 LGVIGAFGAIEBPPLKLLTRPRSSQIPBEGFLTSGGSNYSLARHTCKALTIFLLQPPSP 1189
Db 310 L-----ESPENQEWLSWIGTSQDAVASNIIVAAALVWVYPSV 348
Qy 1190 KLPHPSTIRRTADLIGRGFTWEPYMDVSAVLMGLLELCADAEKOLANITWG-----1242
Db 349 KAKLSLIVWSQLIKLV---MSMNDYSSSTA-----ELLAEGMENTWKAELGAEITHEM 399
Qy 1243 -----LPLSPAADSARSARHALSL-----IATAPPAFITTAKEVHRHT 1282
Db 400 SDILFQIECLSTAPSSNAINKTAVITMRREALVGTLLPSLAMADVTFGFGVIESQIWA-- 457
Qy 1283 ALAANTOSQOQNMHTTTLARAKGBILRVEILIEBKMPDVVDLIVEVMDIIMY-----1334
Db 458 -----TSSDSPVVASIK-----TIIRV-----VRGAPKSLVPLDKAINVILHTMDPSNLI 504
Qy 1335 -----CLEGSIVKKGLOECFPAICRFPMVSYVERNHRIAVG-----ARHGSVALYDIRTG 1385
Db 505 MRKACIISMMALREMARVFP-----MVALNESMTKLAVGDAIGEYDITRVYDIESV 558
Qy 1386 KCOITIGHKGP-----ITAVAPADGRYLATYNTSDSHISFQMNTS 1427
Db 559 TKRIILDAGPPGLPSLLAGSSNRTTTLISVLVSFSPDGEGLVAFSENGLMIRWWSLGS 618
Qy 1428 LLGSIGMNSAPQRCIKTYQVPPVOPASPGS-----1459
Db 619 WHERLS--RSLTIQCTKLIYVPPWEGFSPNSARLSIISILLDDKPRSEKELGEADNLK 676
Qy 1460 ---HNALKLARLIWTSNRNVILMAH 1481
Db 677 LLLHLNLDLSYRLHWIGGKTIKLTTRH 701

RESULT 10

US-10-369-493-20242
; Sequence 20242, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20242
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: No. US20030233675A1toc punctiforme

US-10-369-493-20242

Query Match 2.8%; Score 220.5; DB 15; Length 1005;
Best Local Similarity 20.3%; Pred. No. 1.2e-08;
Matches 124; Conservative 93; Mismatches 251; Indels 143; Gaps 24;
Qy 22 ISAVLLTDDGATIVTGCHDGOICLDWLSVELQINPRALLFGHTASITCLSKACASSDKQY 81
Db 465 IRSIAYSPNGQQLVSASADTKTIKWDVSSGKLLK---TLTGHTSAV---SSVAYNPNGQ 518
Qy 82 IVSASESGMCLWDVSDGRICIEFTKLACTHTGQIQYQFVSGNOREGRLLCHGHYEP-ILV 140
Db 519 LASASDNTIKIWDISSG-----KLLKTLPLGHSSVNVSVAYNPNGQQLASANDKTIKI 572
Qy 141 VDATSLEVLYSAVSKISPDWISSMSIIRSHRTQEDTVVALSVTGILKWIVTSEISDMQD 200
Db 573 WDINSGLKLSLUTGHSS---EVNSVAYSPNGQ---LASASPDNTIKIWDISS--GKLJK 624
Qy 201 TEPIFEESKPIYCONCOSISFCAPTQSRLLVVCVSKYRVFD-----AGDYSLLC 250
Db 625 T---LTGHSNVVP-----SVAYSPNGQHLASASADTKIKIWDVSSGKPLKSLAGHSNVVP 676
Qy 251 S-GPSENGOTWTGDFVSSDKVI-IW-TENGQSIYIKLPASCLPASDPSRSDVGKAVEML 307
Db 677 SVAYSPNGQQLAS---ASDDKTIKWDISNG---KPLESMTDHSRNVSV-----Y 722
Qy 308 IPPVOHILLDRDKELLICPPVTRFYGCREYFHKLLIQGSSGRININWISDTADKQGS 367
Db 723 SPNGQHLASPSYDK-----TIKIWNVS-----744
Qy 368 BEGLAMTTSISLQEAFLKINPCAGIIDQLSVIP-----NSNEPLK-----V 409
Db 745 -SGKLLKTLTGHSSESVNSVAYSPNG--QQLASASWDKTIKWDVNSGKPLKTLIGHSSV 801
Qy 410 TASVYIPAHGRVLCRGEDGSIIVIPATQTAIVOLLQGEHMLRRGPHHTLGRHKVTC 469
Db 802 NSVAYSPNGQQLASASFDNTIKIWDVSSGKLL-----KTLTGHSNAVSS 845
Qy 470 LLYPHQVSARYDORYLISGGVDFSVLIWDIFSGEMKHIPCVHGGBEITOLLVPPENCARV 529
Db 846 VAYSP-----NGQQLASASLONTIKIWDVSSAKLLKTLTGHSDAVSSVAYSPNG-----894
Qy 530 QHCISVASDHSVGLLSLREKKCIIMLASRHLFPFIQVIKWRPDDYLVVGCSDGSVYVWQM 589
Db 895 -QQLASASDNTIKIWDVSSGKLLKSLGHSNAVSVIAYSPNGQQLASASADNTIKIWDV 953
Qy 590 DTGALDRCCVMG 600
Db 954 SSGKLLKSLSG 964

RESULT 11

US-10-369-493-20247
; Sequence 20247, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20247
; LENGTH: 626
; TYPE: PRT
; ORGANISM: No. US20030233675A1toc punctiforme

; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19046
; LENGTH: 1136
; TYPE: PRT
; ORGANISM: Anabaena PCC7120
US-10-369-493-19046

Query Match 2.6%; Score 202; DB 15; Length 1136;
Best Local Similarity 19.3%; Pred. No. 5.7e-07;
Matches 125; Conservative 83; Mismatches 250; Indels 190; Gaps 20;

QY 1 MAGNSVLPIVINGR-----KAPHCISAVLLTDDGATIVTGCHDQICLWDLSVEL 52
DB 580 IASGSLDKTIKLSRDRGLFRTLNGHEDAVYSFSPDGGQTIASGSGDKTIKLW----- 633
QY 53 QINPRALLFGHTASITCLSKACASSDKQYIVSASEGEMCLMDVSDGRCIEFTKLACTHT 112
DB 634 QTSQGLTKITITGHEQTVNNVYSPDGKNLASASSDHSIKLWDTTSG-----QLLWTLT 687
QY 113 GIOYQFVSGNOREGRLCHGHYPEILV-----DATSEVLVSLVSKISPDWISSMIIR 168
DB 688 GHSAGVITVRSPDQGTIAAGSEDKTVKLWHRQDKLLKTLNG-----HQDWNSLSFSP 742
QY 169 SHRTQEDTVVALSVTGLKWIIVTSEISMDQTEPIFEESKDIYQNCQISIFCAFTQR 228
DB 743 DGK-----TLASASADKTIKLMRIA-----DGKLV-----K 768
QY 229 SLVWCKVWRVFDAGDYSLLCSGSPENG-----QWTGDFVSSDKVIWTEEN 277
DB 769 TLKGNDSVWDVNFSSDGAIASASRDNTIKLWNRHGLETFG-----HS 815
QY 278 GOSY-YIKLPASCLPASDSFRSDVGKAVENLIPVQHI----- 314
DB 816 GGVYAVNPLPDSNIIASASLDNTIRLWQRLISPLFVLACNSGVYAVSFLHDGSIITAG 875
QY 315 -----LLDRKDELLICPVTFFYCCREYFHKLIIQDSSGRIN-----IWNISDTA 362
DB 876 ADGNIQLMHSQDGLSLKTLFGNKAIYGI-----SFTPDGDLIASANADKTVKIRVRD-- 928
QY 363 DKQSEGLAMTTSISLQEAFLDNLPCAGIIDQLSVIPNSNEPLKVTASVYIPAHGRVLV 422
DB 929 -----GKALKTLIGHNEVKNVNSPDG-----KTLASASRDN-----TVKLMNVSDGKF- 973
QY 423 CGREDGSIIVPATQTAIVOLLQGEHMLRGWPPHRTLGRHKNKVTCLLYPHOVVSARYDQ 482
DB 974 -----KKTLLKHTDEV-----FWVSFSPDG 993
QY 483 RYLISGVDPSVLIWDIFSGEMKHI FCVHGGEITQLLVPPENCARSVQHICISVASDHSV 542
DB 994 KIITASASADKTIIRLWDSFGSGLNLTSLPAHNDLVYSVNFNPDG-----SMLASTSADKTV 1047
QY 543 GLLSLRKKCMLASRLHFPQIVIKWRPSDDLYLVGCSGGSVVYVQMD 590
DB 1048 KLWRSHDGLLHTFSGHNSVYSSFSFGRIYASASEDKTVKIWIQID 1095

RESULT 14
US-10-369-493-11348
; Sequence 11348, Application US/10369493

; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11348
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Methanosarcina mazei
US-10-369-493-11348

Query Match 2.5%; Score 197.5; DB 15; Length 429;
Best Local Similarity 18.2%; Pred. No. 2.6e-07;
Matches 104; Conservative 85; Mismatches 172; Indels 209; Gaps 21;

QY 22 ISAVLLTDDGATIVTGCHDQICLWDLSVELQINPRALLFGHTASITCLSKACASSDKQY 81
DB 28 VFSIAVTPNGKAVSGSHDGTAKVMDLEKWRIRS---LRAHSKSIATAFA---ITSQGL 81
QY 82 IVSASESGMCLWDVSDGRCIEFTKLACTHTGTFYQFVSGVGNQREGRLCHGHYPEILVV 141
DB 82 VVLGSLDGNLWVNWLETGE-----EKAA-----FKEHS----- 109
QY 142 DATSEVLVSLVSKISPDWISSMIIRSHRTQEDTVVALSVTGLKWIIVTSEISMDQOT 201
DB 110 -----EPITEIV--ITPD-----GKRAVSGSDNTLKW---DLEKMEEL 144
QY 202 EPIFEESKPIYQNCQISIFCAFTQSRLLVVCVKWRVFDAGDYSLLCSGSPENG-QTW 260
DB 145 TTLISHSN-----SVSKIAIT-----PSGKYAI--SGSDNTLKW 178
QY 261 TGGDFVSSDKVIWTEGQSYIYKLPASCLPASDSFRSDVGKAVENLIPVQHIILDRKD 320
DB 179 ---DLKCLDEETISGHSKV----- 196
QY 321 KELLICPPVTRFPYCCREYFHKLIIQDSSGRININWISDTADKQSEGLAMTTSISLQ 380
DB 197 NKIVITPD-----GKLAVSSSYDGTAKVMDLKTKEK----- 228
QY 381 EAFDKLNPACAGIIDQLSVIPNSNEPLKVTASVYIPAHGRVCGREDGSIIVPATQTAI 440
DB 229 -----VTLKGHSGP--VTDVFTPDGRIISGSDDKTLRWD----- 263
QY 441 VOLLOGEHMLRGWPPHRTLGRHKNKVTCLLYPHOVVSARYDQYLISSGVDFSVIWDIF 500
DB 264 -----LKKG---NMTLKGKREVT-----SVAITSDGKYAISGSDFTIKVMDLE 305
QY 501 SEMKHI FCVHGGEITQLLVPPENCARSVQHICISVASDHSVGLLSLRKKCMLASRLH 560
DB 306 NGKIKVTLGHEKNYISTISIIPN-----KNCIVSSSHDETLLKVMWDLDRGIDTITLIGHS 359
QY 561 PFIQVWKWRPSDDLYLVGCSGGSVVYVQMD 590
DB 360 GSVSSVAITPDGKSIIVSASGSGTHKIWSLE 389

RESULT 15
US-10-369-493-18905
; Sequence 18905, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.


```

; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18905
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Anabaena PCC7120
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(478)
; OTHER INFORMATION: unsure at all xaa locations
US-10-369-493-18905

```

Search completed: June 30, 2005, 17:28:42
Job time : 202 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 17:12:56 ; Search time 55 Seconds
(without alignments)
2606.600 Million cell updates/sec

Title: US-10-645-335-2
Perfect score: 7773
Sequence: 1 MAGNSLVLPVILWGRKAPTH.....TSNRNVILMAHDGKEHRFMV 1490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 791: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6023	77.5	1160	T00272	hypothetical prote
2	518.5	6.7	1446	T13018	hypothetical prote
3	234	3.0	1526	AC2239	WD-40 repeat prote
4	230.5	3.0	1227	AE1810	WD-40 repeat prote
5	229	2.9	1258	AI2155	WD-repeat protei
6	217.5	2.8	1101	T26919	hypothetical prote
7	202	2.6	1683	AF2071	WD-40 repeat prote
8	197.5	2.5	2241	T16064	hypothetical prote
9	192	2.5	1049	T42045	beta transducin-li
10	191	2.5	1151	T33777	hypothetical prote
11	189	2.4	494	T19550	hypothetical prote
12	188.5	2.4	1276	T18526	SREBP cleavage act
13	186	2.4	1356	T18521	beta transducin-li
14	182.5	2.3	1184	A96638	hypothetical prote
15	182	2.3	943	S59317	DIP2 protein - yea
16	181	2.3	1693	S76086	beta transducin-li
17	180.5	2.3	1747	AC1842	WD-40 repeat prote
18	180	2.3	1711	AD1842	WD-40 repeat prote
19	174.5	2.2	579	T22703	hypothetical prote
20	173.5	2.2	1216	H85023	hypothetical prote
21	172	2.2	981	T18234	beta transducin ho
22	170.5	2.2	876	T51507	WD40-repeat protei
23	170	2.2	1189	AI2493	WD-repeat protei
24	167.5	2.2	786	AG2375	WD-40 repeat-prote
25	166	2.1	2629	T30987	telomerase-associa
26	165.5	2.1	1223	T17345	hypothetical prote
27	165	2.1	677	AE1861	serine/threonine k
28	163.5	2.1	961	E86245	hypothetical prote
29	163	2.1	1194	T03818	apoptotic proteina

30	161.5	2.1	701	2	T16607	hypothetical prote
31	160	2.1	504	2	T50983	probable pleiotrop
32	160	2.1	1189	2	AH2154	WD-repeat protein
33	156.5	2.0	530	2	T20360	hypothetical prote
34	156.5	2.0	571	2	T20359	hypothetical prote
35	153	2.0	765	2	T49346	conserved hypotet
36	152.5	2.0	1016	2	T19006	ankyrin related pr
37	152	2.0	788	2	S53923	probable membrane
38	151	1.9	515	2	S19487	hypothetical prote
39	151	1.9	1215	2	I52882	autoantigen - huma
40	150.5	1.9	499	2	A56021	probable cell divi
41	149.5	1.9	605	2	T38932	probable sulfur me
42	149	1.9	473	2	T33805	hypothetical prote
43	149	1.9	760	2	T41644	hypothetical trp-a
44	148	1.9	486	2	S49820	PRLL protein - Ara
45	146.5	1.9	589	2	AG2400	WD-repeat protein

ALIGNMENTS

RESULT 1

T00272
hypothetical protein KIAA0541 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 09-Jul-2004
C:Accession: T00272
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: Z14086; MUID:98290545; PMID:9628581
A:Accession: T00272
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1160 <NAG>
A:Cross-references: UNIPROT:Q9Y4E6; EMBL:AB011113; NID:g3043605; PIDN:BA25467.1; PID:g
A:Experimental source: brain
C:Genetics:
A>Note: KIAA0541

Query Match		77.5%; Score 6023; DB 2; Length 1160;
Best Local Similarity		100.0%; Pred. No. 0;
Matches 1160; Conservative		0; Mismatches 0; Indels 0; Gaps 0;
QY	331	RFYGCREFYFHKLLIQGSSGRINIWNISDTADKQGSSEGLAMTTTISIQEAFDKLNPCP 390
Db	1	RFYGCREFYFHKLLIQGSSGRINIWNISDTADKQGSSEGLAMTTTISIQEAFDKLNPCP 60
QY	391	AGIIDLSPVPSNEPLKVTASVYIPAHGRLVCGREDGSIIVPATQTAIVQLLQGEHML 450
Db	61	AGIIDLSPVPSNEPLKVTASVYIPAHGRLVCGREDGSIIVPATQTAIVQLLQGEHML 120
QY	451	RGWPPHRTLGRHNKVTCLLYPHQVSARYDORYLISGGVDFSVIITWDFSGBMKHFV 510
Db	121	RGWPPHRTLGRHNKVTCLLYPHQVSARYDORYLISGGVDFSVIITWDFSGBMKHFV 180
QY	511	HGGEITQLLVPPENCARSVQHICISVADSHSVGLSLREKCKICMLASRLHFPITQVIKWRP 570
Db	181	HGGEITQLLVPPENCARSVQHICISVADSHSVGLSLREKCKICMLASRLHFPITQVIKWRP 240
QY	571	SDYLVVGSGDSGVYVWMDTGALDRCVNGITAVEILNACDEAVPAVDSLSHPANLKQ 630
Db	241	SDYLVVGSGDSGVYVWMDTGALDRCVNGITAVEILNACDEAVPAVDSLSHPANLKQ 300
QY	631	ANTRSLAALKMAHKKLQTLATNLLASEADPKGNLPKYSNLSLMVQAATKTNLTDPDIHV 690
Db	301	ANTRSLAALKMAHKKLQTLATNLLASEADPKGNLPKYSNLSLMVQAATKTNLTDPDIHV 360
QY	691	LFDFVEALIIQLLTEASRPNTALISPENLQKAGSSDKGGSFLLTGKRAAVLFQQVKETI 750
Db	361	LFDFVEALIIQLLTEASRPNTALISPENLQKAGSSDKGGSFLLTGKRAAVLFQQVKETI 420
QY	751	KENIKEHLDDDEEDSEINRQRRESDPYRSSKSKPLTLLEYNLTMTAKLFMSCLHAW 810

Db 421 KENIKGHLDDDEEBEIMRQREESDPPEYRSSKSKPLTLLEYNLTDWAKLFMSCLHAW 480
Qy 811 GLNEVLDEVCLDLRLGMLKPHCTVTSFGLSRGGHMSLMLPGYNOPACKLSHGKTEVGRKLP 870
Db 481 GLNEVLDEVCLDLRLGMLKPHCTVTSFGLSRGGHMSLMLPGYNOPACKLSHGKTEVGRKLP 540
Qy 871 ASGVGKGTGYGSRVATTQHLISIIISLANTLMSMTNATFTIGDHMKKGTPRPPRPSTPDL 930
Db 541 ASGVGKGTGYGSRVATTQHLISIIISLANTLMSMTNATFTIGDHMKKGTPRPPRPSTPDL 600
Qy 931 KARGSPPTSSNIVOGQIKQVAAPVWSARSADHSGSDPPSAPALHTCFLVNEGWSQLAAM 990
Db 601 KARGSPPTSSNIVOGQIKQVAAPVWSARSADHSGSDPPSAPALHTCFLVNEGWSQLAAM 660
Qy 991 HCVMPLDLLGLDKFRPPLLEMLARRWQDRCLVREAAQALLLAELRRIEQAQKEAIDAW 1050
Db 661 HCVMPLDLLGLDKFRPPLLEMLARRWQDRCLVREAAQALLLAELRRIEQAQKEAIDAW 720
Qy 1051 APYLPOVIDHVISPGVTSEAAQTITAPDASGPEAKVQBEHDLVDDDDITTCGLSSVPQM 1110
Db 721 APYLPOVIDHVISPGVTSEAAQTITAPDASGPEAKVQBEHDLVDDDDITTCGLSSVPQM 780
Qy 1111 KKISTSYEERRKQATAIVLLGVIGABFGABIEPPKLLTRPRSSQIPEGFGLTSGGSNYS 1170
Db 781 KKISTSYEERRKQATAIVLLGVIGABFGABIEPPKLLTRPRSSQIPEGFGLTSGGSNYS 840
Qy 1171 LARHTKALTFLILLOPPSPKLPKPHSTIRTPAIDILIGRFTVWEPYMDVSAVLMGLLELCA 1230
Db 841 LARHTKALTFLILLOPPSPKLPKPHSTIRTPAIDILIGRFTVWEPYMDVSAVLMGLLELCA 900
Qy 1231 DAEKOLANITWGLPLSPAADSARSARHLSLIATARPPAFITTIKAEVHRHTALAAANTOS 1290
Db 901 DAEKOLANITWGLPLSPAADSARSARHLSLIATARPPAFITTIKAEVHRHTALAAANTOS 960
Qy 1291 QQNQHMTTLARAKGEILRVIEILLIEKMPDTPVLLVEVMDIIMYCLEGLSVKKGLOECF 1350
Db 961 QQNQHMTTLARAKGEILRVIEILLIEKMPDTPVLLVEVMDIIMYCLEGLSVKKGLOECF 1020
Qy 1351 PAICRFPMWSYERNHRIAGABHGSVALYDITGKCQTHGHKGPITAVAFAPDGRYLA 1410
Db 1021 PAICRFPMWSYERNHRIAGABHGSVALYDITGKCQTHGHKGPITAVAFAPDGRYLA 1080
Qy 1411 TYSNTDISHISFQWNTSLGSGIMLNSAPOLRCIKTVQVPPVPQSPGSHNALKLARLIW 1470
Db 1081 TYSNTDISHISFQWNTSLGSGIMLNSAPOLRCIKTVQVPPVPQSPGSHNALKLARLIW 1140
Qy 1471 TSNRNVLMAHDGKEHRFMV 1490
Db 1141 TSNRNVLMAHDGKEHRFMV 1160
RESULT 2
T13018
hypothetical protein F8L21.60 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
R:Accession: T13018
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirks, W.; Stiekema, W.; Bancroft, I.; Mew
submitted to the Protein Sequence Database, July 1999
A:Reference number: Z17587
A:Accession: T13018
A:Molecule type: DNA
A:Residues: 1-1446 <BEV>
A:Cross-references: UNIPROT:Q9SUT4; EMBL:AL096882; GSPDB:GN00062; ATSP:F8L21.60
A:Experimental source: cultivar Columbia; BAC clone F8L21
C:Genetics:
A:Gene: ATSP:F8L21.60
A:Map position: 4
A:Introns: 895/3; 930/2; 953/3; 1221/3; 1289/2; 1330/3; 1399/3
Query Match 6.7%; Score 518.5; DB 2; Length 1446;
Best Local Similarity 19.9%; Pred. No. 1.3e-27;

Matches 132; Conservative 100; Mismatches 228; Indels 196; Gaps 29;

QY 912 DHMKGTRPRSTPDLKARGSPPTSSNIVGQIKQVAAFPVVSARSADHSGSDPPSA 971
|||:
Db 926 -----ATYSSNLA----- 933
QY 972 PALHTCFVNEGSQLAAMHCVMPLDGLLDKFRPPLEMLARRWDRCLVEEAAQALL 1031
|||:
Db 934 -AYYTRNLAEK-----YDDL-----KPLLQLLVTFWQDSEQVRAARSUF 974
QY 1032 --LAEARRIQAGRKEAIDAWAPYLPQYIDHVISPGVTSEAAQTITAPDASPEAKVQE 1089
|||:
Db 975 HHTASL-----AIPPLCSDH-----ASEHAELVRSLSGISLNEPKVLS 1013
QY 1090 EHDLDVDDITTTGCLAS--VPQMKKISTSYEERRKQATALVLLGVTAEPGABIRPPKLL 1147
|||:
Db 1014 ----TGEYEPTNSLSEHIOARLSQABE-----SEILSWLE----- 1047
QY 1148 TRPRSSQIPEGFGLTSGGNSYLSARHTCKALTF-----LLLOPPSPKLPPTHSTIR----- 1198
|||:
Db 1048 ----SPEMDWTISCVCGETSQDAAAHIIIVAAALSIWPSLVKPLGLAMLVVHKLLNLVMA 1102
QY 1199 -----RTAIDLICRGF-TWPEPYM--DVSALVLMGLLELCADAEKQLANITMGL-----PLS 1246
|||:
Db 1103 MSEKYSTAAELLSEGMEITTKTWIGDIPRIV-----SDIFFQIECVSSSVGAYQVV 1155
QY 1247 PAADSARSARHALSLIATARPPAFITIAKEVHRHTALAAANTOSQONMHTTTLARAKEI 1306
|||:
Db 1156 PSSIKETVELVPLSLAMADVGLFSIIESQIW-----STASDSPVHVVSILR-----TL 1204
QY 1307 LRVE-----TLIEKMPDVVDLLVEVD-----IIMYCLEGSLVKKGLQECRPAIC 1354
|||:
Db 1205 IRIIRAAPRNLVHLEK-----AVNFVLQTMPSNTVMKTKLQTSMATLREVVRVFP----- 1257
QY 1355 RFYVMSYERNHRIAVG-----ARHGSVALYDRTGKQTIHGHKGP----- 1396
|||:
Db 1258 ----MVLINDSSTRLAIGDVTETNNACIHIYDWRMTKIRVLDASGPPGLPNPLRGASES 1314
QY 1397 ----ITAVAPADPGRYLATYSNTDHSIFWQMTSLGSI-----GMLNSAPQLRCIKTY- 1447
|||:
Db 1315 AVTTAISALSFSPDGEGLVAFSEGLMIRWWS-----LGSVWWEKLSQSITPQCTKLIF 1369
QY 1448 -----QVPPVQSPASGSHNALKLARLI-----WTSN 1473
Db 1370 IHPWDGFSSSSRSTSVISSINDEQELPQETAKNISHVERLQVLQVHLDLSYRLEWASE 1429
QY 1474 RNVILMAHD 1482
Db 1430 RKVVLTRHN 1438

RESULT 3
AC2239
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AC2239
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2239
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1526 <KUR>
A:Cross-references: UNIPROT:Q8YR11; GB:BA000019; PIDN:BA075165.1; PID:g17132599; GSPDB:G
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3466

Query Match 3.0%; Score 234; DB 2; Length 1526;
Best Local Similarity 20.1%; Pred. No. 1.8e-07;

Matches 132; Conservative 100; Mismatches 228; Indels 196; Gaps 29;

QY 19 THCISAVLLTDDGATIVTGDHQCICLWDLVSLVELQINPRALLFGHTASITCLSKACASSD 78
|||:
Db 1032 TSCVRSVFPSSDGMALASGDDQTVRLMDIS---SGNCLVTLQGHT---SCVRSVVFSPD 1085
QY 79 KQYIVSASSGEMCLWDVSDGRCIBFTKLACTHTGTQIQYQFVSGNOREGRLCHGYPEI 138
|||:
Db 1086 GAMLASGDDQIVRLWDISSGNCL-YTLQGYT-SWRFVLFVS-----PNGVTTLANGSSDOI 1139
QY 139 L-VVDATSLVLSVLSKISPDWISSMSIIRSHRTQEDTVWALSVTG---ILKVIWVITSEI 195
|||:
Db 1140 VRLWDISSKCLYTLQG---HTWVNAVAF-----SPDATLASGSDQTVRLWDISS--- 1189
QY 196 SDMQDTEPIFEBESKPIYQNCQSIISFCAPTQSLVAVCSKYW---RVFDAGDYSLCSCG 252
|||:
Db 1190 ----SKCLY-----ILQHTSWVNSVWFNP-DGSTLASG 1218
QY 253 PSBNGQWTGGDFVSDKVI-TWTENGQSYIYKLPASCLPASDSFSDYGVKAVENLIPPV 311
|||:
Db 1219 ----SSDQTVRLWEINS-----SKCLCTFGHTSWNSVVFN----- 1251
QY 312 QHILLDRDKELLICPPVTRFFVGCREYFHKLLIQDSSGRNLNIWNISDTADKQSGEGL 371
|||:
Db 1252 ----PD-----GSMLASGSDKTVRLWDISSKCLHTFGHT 1284
QY 372 AMTTSISLQEAFLKLP---CPAGIIDQ---LSVIPNS-----NEELKVYASVYIPA 417
|||:
Db 1285 NVNNSVA-----FNPDSGMLASGSDQTVRLWEISSKCLHTFGHTSWNSVVFSPD 1337
QY 418 HGRVLCGRDGGIVIVPATQTAIVQLQGEHMLRQWPPHRTLRGRNKVTCLLYPHQVS 477
|||:
Db 1338 GTMLASGSDQTV-----RLWSISSGECLY-----TFLGHTNVGVIFSP--- 1378
QY 478 ARYDQRYLISGGVDFSVIIVDFSGEMKHFVCHGGEITQLLVPPNCSARVQHCICSA 537
|||:
Db 1379 ----DGAILASGSDQTVRLWSISSGKCLYTLQGHNNWVGSIVFSPDGT-----LIASGS 1429
QY 538 SDHSVGLLSLRKKCKIMLASRLFPQIVIKWRPDDYLVVGGSDGVSYYVQMGTGALDRC 597
|||:
Db 1430 DQOTVRLWNISGECLYTLHGHSVRSVAPSSDGLILASGSDDETIKLWDVKTG---EC 1486
QY 598 V-----MGITAVELINACDEAVPAVDLSLHAPVNLKQAMTRSLAALKN 642
|||:
Db 1487 IKTLRSEKIVYEGNITSVRGLTEVERA-----TLKTLGAVEN 1523

RESULT 4
AE1810
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AE1810
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE1810
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1227 <KUR>
A:Cross-references: UNIPROT:Q8Z0R1; GB:BA000019; PIDN:BA077553.1; PID:g17135007; GSPDB:G
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0029

Query Match 3.0%; Score 230.5; DB 2; Length 1227;
Best Local Similarity 19.6%; Pred. No. 2.2e-07;
Matches 142; Conservative 102; Mismatches 248; Indels 231; Gaps 31;

QY 22 ISAVLLTDDGATIVTGDHQCICLWDLVSLVELQINPRALLFGHTASITCLSKACASSDKQY 81

Db 605 VSVKSPDKGYATGLMGEIRLWQTSNKKQLR---IYKGTAWWAF---FSPDSRM 658
QY 82 IVSASEGEMCLWDVSDGRCIEFTKLAHTGTGTFYQFVSGNOREGLLCHGYPE-ILV 140
Db 659 LASGSADSTIKLMDVHTGECL---KTLTKNTN---KVYSVAFSPDGRILASASQDTIKL 712
QY 141 VDATSLEVLVSKISPDWISSMSIIRSHRTQEDTVVALSVTGILKVIIV-----191
Db 713 WDIATGNCQOTLIG--HDDWVWVTFSPVTDRLPLLLASSADQHKLMADVATGKCLKTL 770
QY 192 ---TSET-----SDQDTE-----PIFEESKPIY---CONCOS 219
Db 771 KGTREHVSVPFSDGQTLASSGDESTVRLWDVKTGCQWQIFGSHKKYISVRFSPDGT 830
QY 220 ISPCAFQTRSLV-----VC-----SKYWRV-FDAGDYSLLCSPSENGQTTWGD 264
Db 831 LASCG-EDRSIKLWDIQRGECVNTLWGHSSQVWAIAPSPDRTLISGDDQTLARLW---D 886
QY 265 FVSSDKVII---WTENGQSYIYKLPASCLPASDSFRSDVGVKAVENLIPVQHILLDRDK 321
Db 887 VITGNSLNILRGYTRDYSVAFS-PDSQILASG--RDDYITGLWNLKTGECHPLRGHQR 943
QY 322 ELLICPEVTRFFYGCREYFH---KLIIQGDSSGRLNWNISDTADKQGSBEGLAMTTSIS 378
Db 944 IRSVA-----FHPDGKILASGADNTIKLWDIST-----973
QY 379 LQAFDKLPCPAGIIDQLSVIPNSNEPLK-----VTASVYIIPAHGRLVCGREDGSI 430
Db 974 -----NHSKYIRTLTGHTNWMVWVVFSPDKHTLASSSEDRTI 1010
QY 431 VIYPATQTAIVQLLOGEHLMLRGWPH-----RTLGRHKNKVTCLLYPHQVSARYDQYL 485
Db 1011 -----RLWDXDTGDCQLKJGHSHWVWVAFSP-----DGRIL 1043
QY 486 ISGVDPSVLIIDIFSGEMKHIFCVHGEITQLLVPENCARSVQH-----CICSVASDH 540
Db 1044 ASSADSEIKIWDVASK-----CL-----QTLTDPQMWISVAFSLDGTLLASASEQ 1092
QY 541 SVGLLSRLREKKICIMLASRHLFPQIVIKWRPSDDYLVVGCSGDSVYVQMDGTALDRCVMG 600
Db 1093 TVKMLNKTGECVHTLKGEHKQVYSVAFSPNGQIAASGSEDTTVKLWDISTG-----1144
QY 601 ITAVEILNACDEAVPAVDLSHPAVNLKQAMTRRSIALKNAHKLQTLATNILLASEA 660
Db 1145 -----SCVDTLKH-----GHTAAIRSVAF-----SPDGRLLASGS 1174
QY 661 SDK 663
Db 1175 EDE 1177

RESULT 5
A12155
WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: A12155
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
DNA Res. 9, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: A12155
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1258 <UNP>
A:Cross-references: UNIPROT:Q8VTC2; GB:BA000019; PIDN:BAB74499.1; PID:g17131893; GSPDB:G
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2800

Query Match 2.9%; Score 229; DB 2; Length 1258;
Best Local Similarity 22.5%; Pred. No. 2.9e-07;
Matches 137; Conservative 93; Mismatches 208; Indels 172; Gaps 34;
QY 20 HCISAVLLPDDGATVWGCHDQICLWLSVLEQLINPRALLFGHTASITCLSKACASSDK 79
Db 727 HEVFSVAFPHDGETLASAGDKTIKLDIQDGTCLQ---TLTGHTDWVRCVA---FSPDG 780
QY 80 QIVSASEGEMCLWDVSDGRCIEFTKLAHTGTGTFYQFVSGNOREGLLCHG-----133
Db 781 NTLASSAADHTIKLMDVSGKCLRTLK---SHTG---WVRSVAFSADGQTLASGSDRTI 834
QY 134 ----HYPEILVVDATSLVLSKISPDWISSMSIIRSHRTQEDTVVALSVTGILKVI 188
Db 835 KIWVHTGECLTYTGHTNSVISIA--YSPD-----SKILVSGSGDRTIKL 878
QY 189 WIV-----TSEISDMQDTEPIFEESKPIYQNC-OSISF--CAFTQSRSLVV 233
Db 879 WDCQTHICIKTLGHTNECVS-----AFSPDQTLACVSLDQSVRLMNCRTGQ-----927
QY 234 CSKYW-----RVFDAGDYSLLCSPSENGQT-----WTGDDFVSSDKVILWTENGOS 280
Db 928 CLKAWYGNTDWALPVAFSPDQILASG--SNDKTVKLMWQTKYISS-----LEGHTD 979
QY 281 YIYKL---PASCULPASDSFRSDV-----GRAVENLIPVQHILLDRKDKELLICPPVT 330
Db 980 FIYGIAFSPDSQTLASASTDSSVRLWNISTGOCFO-----ILLEHTD-----1021
QY 331 RFFYGCREYFH---KLIIQGDSSGRLNWNISDTADK-----SPTADK-----QGSBEGLAMT 374
Db 1022 -WYAV--VFHPQGIATGASADCTVKLWNISTGQCLTSLSEHSDKILGMAWSPDG-QLL 1077
QY 375 TSISLQEAQKLNPCPAGIIDQLSVIPNSNEPLKVTASVYIIPAHGRLVCGREDGSIIVVP 434
Db 1078 ASASADQSVRLWDCCTGRCVGILR--GHSN---RVYSAFSP-----NGEIIATC 1122
QY 435 AT-QTA-IVQLLOGEHLMLRGWPHRTLGRHKNKVTCLLYPHQVSARYDQYLISGVDF 492
Db 1123 STDQTVKIWDWQGGKCL-----KTLTGHTNWV-----FDIAFSPDGKILASASHDQ 1168
QY 493 SVIIMWIDIFSGEMKHIFCVHGEITQLLVPENCARSVQHICVSAADSHSVGLLSREKKC 552
Db 1169 TVRIWDVNTGKCHICIGHTHLVSSVAFSPDG-----EVAASGSDQTVRIWNVKTGEC 1222
QY 553 I-MLASRHLF 561
Db 1223 LQILRAKRLY 1232

RESULT 6

T26919
hypothetical protein Y45F10B.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T26919
R:McMurray, A.
submitted to the EMBL Data Library, January 1998
A:Reference number: Z20286
A:Accession: T26919
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1101 <WIL>
A:Cross-references: UNIPROT:Q62471; EMBL:AL021487; PIDN:CAAL16357.1; GSPDB:GN00022; CESP
A:Experimental source: clone Y45F10B
C:Genetics:
A:Gene: CBSP:Y45F10B.10
A:Map position: 4
A:Introns: 217/3; 292/1; 451/3; 556/2; 637/3; 755/3; 823/2; 946/1; 1078/3

Query Match 2.8%; Score 217.5; DB 2; Length 1101;
Best Local Similarity 18.0%; Pred. No. 1.5e-06;
Matches 149; Conservative 110; Mismatches 246; Indels 325; Gaps 35;

QY 16 KAPTHCISAVLLTDDGATIVTGCCHDQICLWLSVQLINPRALLFGHTASITCLSKACA 75
Db 417 KGHTAAVYTCCLSSNDLSFLVSTSPDKTVNVMVFS---QSTPTMSLTHHTAKVTC---AIL 470
QY 76 SSDKOYIVSAS-----ESGEMC-----L 93
Db 471 TSDQOYLITASADSSAKMIKLEGEVMSFNDHTGVSVLSQTSNNQFLITGSGDFVQM 530
QY 94 WDVSQDGRGIE-----FTKLACTHTG:QFYQFSVGNQREGRLLCHGH 134
Db 531 WDVNTGKICRMGLMAPVSTLAITSNDFAFWVACEDETLKVSF-TVGQQLHELMHGEH 589
QY 135 YPELLVVDATSLVLYLSVSKI-----SPDWISSMSIIRSHRTQBDTV 177
Db 590 KMSLVCAQDDCOLFAATKSKFCYD:IHNGQIMDLVLTQAQFPFICSLKI-----SSDNY 643
QY 178 VALSVTC-ILKMWIVTSEISDMOD-----TEPIFEESKPIYC-----Q 215
Db 644 FLISPCGPKVTIWNVTKRNDHADVHADKEGFLTAVALSNDKDYAACGTNGGIVALWDL 703
QY 216 NCQISISFCAPT---QRSLLVVCSKYMRVFDAGDYSLLCSG-----PSENG----- 257
Db 704 VCO---CVFTTIQKGDPTICRY-----SVDSQYCSIGNQAGCILILDAQNGGVREL 754
QY 258 -----QWTGCGFVSDD---KVIWTEGQSYIYKLPASCLPASDSFRSDVGKAV 304
Db 755 FMHSSEVLSIMSIVHNKMSICDIQGRMVIWELFGDDTPEMVAI----- 798
QY 305 ENLIPPVHILLDRKOKELICPPVTRFFYGCREFYFKLLIQDSSGRLNIWNISDTADK 364
Db 799 -GVKPPI-----FVPTGRIMVG-----HCSL-----SNKEMKIAP----- 829
QY 365 QGSEEGAMTTSISLOBAFDPKPCPAGIIDLQSVIPNSNEPLKVTASVYIPAHGRILVC- 423
Db 830 --PEEGPPVTRA-----KLS-----HSDS-----ITCFATSPKGFNIAT 862
QY 424 GREDGSIIVPATQTAIVQLQGEHMLRRGWPHRTLRHNRKVTCLLYPHQVSARYDQR 483
Db 863 GSRDMSLSKIWDIKGFLTQVLV-----GHENVVTC-----CCISFDER 900
QY 484 YLISGGVDFSVIWDIFSGEMKHIFCVHGGEITQLLVPENCARSARVQHICSVASDHSVG 543
Db 901 LVVSGARDEKILVWNVQSGDM-----VCTVNTTAAIT 932
QY 544 LLSLRKKKICMLASRHLFP:QVIKWRPDDYLAVGCSGDSGVVYVQMDTGAL----- 594
Db 933 SLSTGDSVTVFST-----TEDGWVETWSTTKGRLSTFNAHRP 971
QY 595 -DRCVMGITAVEILNACDEAVPAADVSLSH-PAVNLQAMTRRSALAKNMMAHKKLQTLA 652
Db 972 IKKLINSYSHRMLLLLENCAQLPILCLHNTPAVGVE--ATRRRSARAQSVSSASNEPVA 1029
QY 653 TN-----LLAS-----EASDGNLPKYSNLSMWQAIKNTLTDPD 687
Db 1030 STSAGEIKKDPILSSNNGNAQAPRATAPKPTFDMLERSKSRSTLIEKD 1079

RESULT 7
AF2071
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AF2071
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2071
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1683 <KUR>

A:Cross-references: UNIPROT:Q8YV57; GB:BA000019; PIDN:BA073823.1; PID:g17131215; GSPDB:
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all12124

Query Match 2.6%; Score 202; DB 2; Length 1683;
Best Local Similarity 19.3%; Pred. No. 3.8e-05;
Matches 125; Conservative 83; Mismatches 25; Indels 190; Gaps 20;

QY 1 MAGNSLVLPVLMGR-----KAPTHCISAVLLTDDGATIVTGCCHDQICLWLSVQL 52
Db 1087 IASGSLDKTIKLSRDRGLFRTLNHGEDAVYVSFSPDQGTIASGSDKTIKLM----- 1140
QY 53 QINPRALLFGHTASITCLSKACASSDKQYIVSASEGEMCLMWDSQRCIEFTKLACTHT 112
Db 1141 QTSDDGTLTKITTGHEQTVANNVYSPDGKNLASASSDHSIKLMDWTTSG-----QLLMTLT 1194
QY 113 GIOFYQFSVGNQREGRLLCHGHYPEILV-----DATSLEVLYLSVSKISPDWISSMSIIR 168
Db 1195 GHSAGVITVRFSPDGTIAAGSEDKTVKLMWRHQDQKLLKTLNG-----HQDWNSLSFSP 1249
QY 169 SHRTQBDTVVALSVTGIKMWIVTSEISDMQDTEPIFEESKPIYCQNCOSISFCAPTQR 228
Db 1250 DGK---TLASASADKTIKLMRIA-----DGKLV-----K 1275
QY 229 SLVAVCSKYMRVFDAGDYSLLCSGSENG-----QWTGCGFVSDDKVIWNTEN 277
Db 1276 TLKGHNSVMDVNFSSDGRKAIASASRDNTIKLWNRHGLELETPG-----HS 1322
QY 278 GQSY-IYKLPASCLPASDSFRSDVGKAVENLIPVQHI----- 314
Db 1323 GGYAVNFLPDSNIIASASLDNTIRLQRPISLPLEVLAGNSGVYAVSFLHDGSIITATAG 1382
QY 315 -----LLRDKKELICPPVTRFFYGCREFYFKLLIQDSSGRLN-----IWNISDTA 362
Db 1383 ADGNIQWISQDGSLLKTLPGNKAIYGI-----SFTQGDLIASANADKTVKLWRVD-- 1435
QY 363 DKQSGEGLAMTTSISLOBAFDPKPCPAGIIDLQSVIPNSNEPLKVTASVYIPAHGRLV 422
Db 1436 -----GKALKTLIGHNEVKNVNFSPDG--KTLASASRDN-----TVKLMVNSDGKP- 1480
QY 423 CGREDGSIIVPATQTAIVQLQGEHMLRRGWPHRTLRHNRKVTCLLYPHQVSARYDQ 482
Db 1481 -----KTKLKGHTDEV-----FWVSFSPDG 1500
QY 483 RYLISGGVDFSVIWDIFSGEMKHIFCVHGGEITQLLVPENCARSARVQHICSVASDHSV 542
Db 1501 KIITASADKTIKLMWDSFGNLIKSLPAHNDLVYSVNFPDQ-----SMLASTSADKTV 1554
QY 543 GLLSLRKKKICMLASRHLFP:QVIKWRPDDYLAVGCSGDSGVVYVQMD 590
Db 1555 KLWRSHDGHLLHTFSGHNSVYVSSSPDGRYIASASEDKTVKIWDID 1602

RESULT 8
TI6064
hypothetical protein F13H8.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: TI6064
R:Ding, H.
submitted to the EMBL Data Library, July 1995
A:Description: The sequence of C. elegans cosmid F13H8.
A:Reference number: Z18455
A:Accession: TI6064
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2241 <DIN>
A:Cross-references: EMBL:U23139; NID:g722370; PID:g722372; PIDN:AAC46678.1; CESP:F13H8.
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F13H8.2
A:Introns: 52/3; 178/2; 269/3; 336/2; 486/3; 701/3; 862/2; 938/3; 1033/1; 1074/2; 1117/

C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33777
R;Ryan, E.; Delehaanty, A.
submitted to the EMBL Data Library, October 1998
A;Description: The sequence of C. elegans cosmid H24G06.
A;Reference number: 221406
A;Accession: T33777
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1151 <RYA>
A;Cross-references: EMBL:AF101309; PIDN:AA69213.1; GSPDB:GN00023; CESP:H24G06.1
A;Experimental source: strain Bristol N2; clone H24G06
C;Genetics:
A;Map position: 5
A;Introns: 18/3; 91/3; 220/3; 353/2; 441/3; 470/1; 510/2; 593/3; 773/2; 825/1; 852/2; 87

Query Match 2.5%; Score 191; DB 2; Length 1151;
Best Local Similarity 18.4%; Pred. No. 0.00012;
Matches 218; Conservative 163; Mismatches 376; Indels 426; Gaps 59;

QY 15 RKAPTHCISAVL---TDDGATVTC-HDQICLW-----DLSVELOINP-RALLFGH 63
DB 129 RSMQSHVSVITVRPTDDSTLISGCHDAQIFTDWRTGHDLVGKLMSPVNAITISF 188

QY 64 TASITCL-----SKACASDQYI--VSAS--ESGEMCL 93
DB 189 DSSM-CVTGSKSVKYWFLPLASDGMNKRGLTSRAILLADKRNVNFDVAFCDSNRMV 247

QY 94 WDVSDGRCTEFT---KLACTHTGIGVQPSV-----GNQREGRLCHGHYPEILVVD 142
DB 248 TVTSSGEILEFNGKVCVY---QFNFFVQVYKVTWKESGDFKAICIAIRVPEGLVG 304

QY 143 ATSLVLYSLVSKISPDWISSMIIIRSHRQEDTVVALSVTGILKWIYTSB-----ISDM 198
DB 305 CSD-----GLVRFLCLIGDLDLFLADL 326

QY 199 QDTEPIFEESKPIYQNCOSIFCAFTQRLSLVGCYKWRVFDAG--DYSLLCSGSPEN 256
DB 327 APPTHLFQDPSNVYDQOLQNHPEDAIFPEPRCLVASSSSPTFVGVADRSLI----- 379

QY 257 GQTWTGDFVSSDKVIITENGOSYIKLPASCL---PASDSFRSDVGRKAVENLIPVQH 313
DB 380 -----EFARENK--SWSFRASLGHGTGSVNCIEPPSPSSP----- 412

QY 314 ILLDRKDKELLICPPVTRFPYGCREFHKLIIODSSGRNLINWIS-DTADKQSGSEGLA 372
DB 413 -----CLPI-----GFLITGGSDGTIRFWNFGDSEDKK--BIANI 446

QY 373 MTTTSISIQEAFDKLPCPAGIIDQL---SVIPNSNEPLKVTAS---VVIAPHR-LVCG 424
DB 447 LCPSL-LKVTF--LDENPDLLVDKRNLESVGPSENQSNSSGVLCVTRVSHDGRMLIAG 503

QY 425 REDGSIVIPA--TQTAIVOLLOGE----- 449
DB 504 TATGMLYIVDLSPSPDIIDVINAHBDVTSIDFSDBASTSSHHDHPIFLASGGRDRFVH 563

QY 450 LRRCWPPHRT-----LRHRNKVTCLLYPHOVSARYDORYLISGVDFSVIINDIFS- 501
DB 564 FRP--IPYSSQFVHCVAVLGDHQAISIKF-----ASNNGQLHLHYTAASDRSLIWKLSF 617

QY 502 -----GEM--KHI--FCV----- 510
DB 618 SDQCEFRVQMISSVASSIGDMNFIKHVDLFFVAGHDMRLRQPDINGKTVREKGTDDVD 677

QY 511 -HGEITQLAVPPENCARSVQHICSVASDHSVGLLSREKKICMLASRLHLPPIQVIRW 569
DB 678 AHSKILKVAIDHSGSA-----ISVCSDKFVVVTDLRSGVCLAVLCGFGAPATATP- 730

QY 570 PSDDY--LVVCGSDGSVYVQMDTGALDRCMG-----ITAVEILNACDEAVP 615
DB 731 -SDDFNKVVITNGSIFITVQLAKNLTERMISAQVRLMBEVTWRTATPDPSLLSGSETIS 789

QY 616 AAVDSLSPAVNLKQAMTRRSALAKNMAHKLQTLATNLLASEASDKGNLPHYNSLM 675
DB 790 GDSNSFGRP-----LCAPD-----FSGSSASLYSDDDDSTRFSS--- 824

QY 676 VQAIKTNLTDPDIHVLFPDVE-----ALIIQLLTHEASRPNTALISPENLOKASGS 726
DB 825 VRSSRTKRILPNGHGLIGDSSFYARVGDSSFSFAVOSAPAVERTHTNLFSD--QYETDV 882

QY 727 SDKGSFLTGKAAVLFQOVKETIKENIKELHLLDDEEDEI-----MQRREESD 777
DB 883 SETQSDFFVSSR-----KTRLPSEDDQDNLGSAQYLAAPINEDRRSAS 926

QY 778 PE-----YRSKSKPLTLLEY---NLTWDTAK-LFMSCLHAW-----GLNEVLDE 818
DB 927 PSLYPPQDOLRGYQSSKSM-MNLRDVTGGVVRVQAKELMMSQIASQRNSQSGSSHLSS 985

QY 819 VCLDRGLMLK-----PHCTVSFGLLSRGHMS-LMLPGYNQPACKLSHGKT 863
DB 986 T--NTSGRMKGMFQOSNNNDWHSSTVDI-----HVSVMATSMTPPTQHQYGYR 1037

QY 864 EYGRKLPASEG-----VGKTYGVSRVTTQHLLSIIISLANTL--MSMTNATPFIGDHMK 915
DB 1038 DHGTSYPMEDSTHPPPLAPRTTSRVLSTAPSQAALQOIQASSPFRKSDRNSLSRRFLE 1097

QY 916 KGPTRP-----PRPSTDLSKARG--SPPTSSNI 942
DB 1098 NGGAQPKTWSPSALANQAPRRSNSNLFPAATNLEVPPTSTNL 1140

RESULT 11
T19550
Hypothetical protein D1054.15 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T19550; T20311
R;Wilkinson, J.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19140
A;Accession: T19550
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-494 <WIL>
A;Cross-references: UNIPROT:Q18295; EMBL:Z73970; PIDN:CAA98247.1; GSPDB:GN00023; CESP:D1054.15
A;Experimental source: clone C29A12
R;Matthews, P.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19255
A;Accession: T20311
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-494 <WIL>
A;Cross-references: EMBL:Z74030; PIDN:CAA98448.1; GSPDB:GN00023; CESP:D1054.15
A;Experimental source: clone D1054
C;Genetics:
A;Gene: CESP:D1054.15
A;Map position: 5
A;Introns: 22/1; 61/2; 249/3; 475/3
C;Superfamily: PRL1 protein; WD repeat homology

Query Match 2.4%; Score 189; DB 2; Length 494;
Best Local Similarity 24.4%; Pred. No. 4e-05;
Matches 79; Conservative 36; Mismatches 143; Indels 66; Gaps 12;

QY 398 SVIPNSNEPLKVTASVVIAPHRVLCGREDGS-----IVIVPATQTAIVOLLOGEHMLR 451
DB 106 SVIASAGTFLAIT-----DGSCKLV-NQOQSAKSGTLLPLPLGNSS-----KGEDNTT 154

QY 452 RG-----WPHRTLRGRNKNVTCLLYPHQVSARYDORYLISGVDFSV 494
DB 155 RSLSPSKAPMMKPKWHPKLYRVASGHTGV-----RAVDVEPCNWFASGGADRII 208

QY 495 IIVDIPSGEMKHFVCHGGEITQLLVPPENCARSVQHICSVASDHSVGLLSREKKICM 554

Db 209 KIWLASGQLKLSITGHISVRVAVKVSPRH-----PFLFGSGEDQVQKWDLEYNKVR 262
Qy 555 LASRHLPPIQVIKWRSDDDYLVGCSGDSYVYVQMDTGALDRCMVGITAVEILNACDEAV 614
Db 263 HYHGLSVAQALSVPFSLDVLVTCARDSTARVWDMRTKAQVHCFAGHTNTVADVVCQSDV 322
Qy 615 PAAVDSLSHPAVNLKQAMTFRSLAALKNMMAHKLQTLA-----TNLLASEASD---KGN 665
Db 323 PQVITASHDATVRLWDLAAGRSWCTL---THHKKSVRALTIHPRLNMFASASPDNIKQWK 379
Qy 666 LPK-----YSHNSLWVQAIKTN 682
Db 380 LPKGEFQNLGSHNAI-INTLSSN 402

RESULT 12
T18526
SRBP cleavage activating protein - Chinese hamster
N/Alternate names: SCAP protein
C/Species: Cricetus griseus (Chinese hamster)
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: T18526
R/Hua, X.; Nohturfft, A.; Goldstein, J.L.; Brown, M.S.
Cell 87, 415-426, 1996
A/Title: Sterol resistance in CHO cells traced to point mutation in SRBP cleavage-activating protein
A/Reference number: Z18949; MUID:97053781; PMID:8698195
A/Accession: T18526
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1276 <HUA>
A/Cross-references: UNIPROT:P97260; EMBL:U67060; NID:g1675219; PID:g1675220; PIDN:AAB191
A/Experimental source: ovary
C/Genetics:
A/Gene: SCAP
A/Function:
A/Description: regulates cholesterol metabolism by stimulating cleavage of transcription factor
C/Keywords: transmembrane protein

Query Match 2.4%; Score 188.5; DB 2; Length 1276;
Best Local Similarity 21.5%; Pred. No. 0.00021;
Matches 136; Conservative 82; Mismatches 196; Indels 219; Gaps 34;

Qy 51 ELQINPRALLFGHTASITCLSKACASSDKQYIVSASEGEMCLWDVSDGRCIEFTKLACT 110
Db 764 ETEIVP-LVLRGHLMDIECL-----ASDGMLLVSCCLAGQVCVMDAQTDGCL--TRIP-- 813
Qy 111 HTGIQYQFVSGNQRGRLLCHGYPEILLVVDATSLVLYSLVS---KISPDWISSMSII 167
Db 814 -----RPGSRDS---CGG-----GAFQENWENRSLSDGKTSPEEPGESPL 853
Qy 168 RSHRTQEDTVVALSVTGILKMWIVTSEISDMQDTRPIFEESKPIYQNCQSIQSCAFTQ 227
Db 854 R-HRPRGPPQALFGQDPOLITCLIDTNFSLVQLPPEP---TQPRHRAGC----- 899
Qy 228 RSLLVVCSKWRVPDAG-DYSLSCSPSENGQWTGDFVSSDKVLIWENGOSYIYKLP 286
Db 900 -----GRADSGYDFSLVRQ-----VYQEEGLAAV-RMP 928
Qy 287 ASCLPASDS-----FRSDVGKAVENLIPPVQHILLDRKDELICPPVTRFFYGCREFHK 342
Db 929 ALRPPSPGSLPOASQEDGAPEKSGPP-----LAWAPSTAGSIWLSLQGN 975
Qy 343 LLIQGSSGRRLNTWISDTADKGSSE---GLAMTTSISLQEAFLKNCPCPAGIIDQLSV 399
Db 976 LIVVGRSSGRLEWDAIEGVLLCCSNDEVSGITALVFLDRIRVAARLN-----GSLDFSL 1031
Qy 400 IPNSN-EPLKV-----TASVY-----IP-----AHGLVC 423
Db 1032 ETHISLSPQFRTPGRGSPSSSVYSSNTVACHLTHVPCAHQKPYALRAAGRLVT 1091
Qy 424 GREDSIVIVPATQTAIVQLLOGEHMLRRGWPPHRTLGRHNKVTCLLYPHQVSARYDQR 483

Db 1092 GSQDHTLRVRLLEDSCCL-----FTLQHGSAIT-TVYIDQTMV----- 1129
Qy 484 YLISGVDPSVLIWIDIFSG-EMKHIFCVHGHEITOLLVPPENCARSVQHICISVAS- 538
Db 1130 -LASGQDGAICLMDVLTGSRVSHTF-AHRGDTVSL-----TCTTS-----CVISSGLDD 1177
Qy 539 -----DHSVGLLSLREKKCICMLASRHLFPIQ-----VIKWRPSDDYLVVVGCSGDS 583
Db 1178 LINIWRSTGI-----XLYSIQDLGGASLGI-----SDNLLVTG-GQGC 1218
Qy 584 VYVQMDTGALDRCMVGITAVEILNACDEAVPA 616
Db 1219 VSFMDLNYGDLLQTV-----YLGKNSEAQPA 1244

RESULT 13
T18521
beta transducin-like protein - Podospora anserina
C/Species: Podospora anserina
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: T18521
R/Saube, S.; Turcog, B.; Begueret, J.
Gene 162, 135-139, 1995
A/Title: A gene responsible for vegetative incompatibility in the fungus Podospora anserina
A/Reference number: Z18944; MUID:96009891; PMID:7557402
A/Accession: T18521
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1356 <SAU>
A/Cross-references: UNIPROT:Q00808; EMBL:L28125; NID:g607002; PID:g607003; PIDN:AAA85775
C/Genetics:
A/Gene: het-el
A/Introns: 761/3

Query Match 2.4%; Score 186; DB 2; Length 1356;
Best Local Similarity 18.4%; Pred. No. 0.00036;
Matches 113; Conservative 76; Mismatches 220; Indels 206; Gaps 22;

Qy 22 ISAVLLTDGATVGTGCHDGOICLWLSVELQINPRALLFGHTASITCLSKACASSDKQY 81
Db 844 VLSVAFSADGQQRVAGSGDDTKIKIWDTA---SGTQTQTLGHHGGSVMSVA---FSPDRR 897
Qy 82 IVSASEGEMCLWDVSDGRCIEFTKLACTHTG-IQFYQFVSGNQRGRLLCHGYPEILLV 140
Db 898 VASGDDTKIKIWDAAAGTC---TQLEHGGRVQSVAFSPDQR----- 939
Qy 141 VDATSLEVLVSLVSKISPDWISSMSIIRSHRTQEDTVVALSVTGILKMWIVTSEISDMQD 200
Db 940 -----VASGDDTKIKIWDAAAGTC--- 959
Qy 201 TEPIFEESKPIYQNCQSIQSCAFTORSLLVVCCKYWRVFDAGDYSLSCSPSENGQWTW 260
Db 960 TQTLGHHGSSVL-----SVAFSPDQRVAGSGDDTKIKIWDTAAGTCTQTLGHHGGSVW 1013
Qy 261 T-----GDPFVSS---DKVI-IW-TENGOSYIYKLPASCLPASDSFRSDVGKAVENLI-- 308
Db 1014 SVAFSPDQRVAGSGDDTKIKIWDATAG-----TCTQTLGHH-----GQWQSVFVS 1060
Qy 309 PPVQHILLDRKDELICPPV---TRFFYGCREFHKLLIQGSSGRRLNTWISDTADK 364
Db 1061 PDGQRVAGSGDDTKIKIWDVAGSTCTQTLG-----HGDS-----VMSVAFSPDG 1105
Qy 365 QSGSEGLAMTTSISLQEAFLKNCPCPAGIIDQLSVIPNSNEPLKVTASVYIPAHGLVCG 424
Db 1106 Q-----RVASG 1111
Qy 425 REDGSIVIVPATQTAIVQLLOGEHMLRRGWPPHRTLGRHNKVTCLLYPHQVSARYDQR 484
Db 1112 SIDGTIKIWDAAAGTCCTQTLG---HGW-----VHSVAFSPDQR 1149
Qy 485 LISGVDPSVLIWIDIFSGEMKHIFCVHGHEITOLLVPPENCARSVQHICISVASDHSVGL 544
Db 1150 VASGSDGTIKIWDAAAGTCCTQTLGHHGGSVMSVAFSPDG-----QRVAGSDTKIKI 1203

QY 545 LSLREKCIIMASRLHPPIQVIRSDDDYLVVCGSDGVSYYVQMDTGALDRCVMGITAV 604
Db 1204 WDTASGTCTQTLEGGHGWQVAFSPDQQRVASGSDNTIKIWDTAGS-----TCT 1254
QY 605 EILNACDEAVPAVD 619
Db 1255 QTLNVGSTATCLSPD 1269

RESULT 14
A96638
hypoetical protein Flp17.7 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
R:Accession: A96638
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96638
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1194 <STO>
A:Cross-references: UNIPROT:O22725; GB:AE005173; NID:g2443881; PIDN:AAB71474.1; GSPDB:GN
A:Gene: Flp17.7
A:Map position: 1

Query Match 2.3%; Score 182.5; DB 2; Length 1184;
Best Local Similarity 17.2%; Pred. No. 0.0005;
Matches 217; Conservative 180; Mismatches 415; Indels 451; Gaps 54;

QY 342 KLLIQDSSGRINWISDADKQSE-----EGLAMTTSISLQAFDK-----LNPCP 390
Db 29 RLFTGGDDYKYNLMAIGKPTSLMKNDATAFYQSLCGHTSAVDSVAFDSAEVLVLAGAS 88
QY 391 AGHIDQLSVIPNNEPLKV-----TASVYIPAHGRVLCYCGEDGSIVVPATQTA 439
Db 89 SGVILKWDV-----EAKNVRAFTGHRNSCSAVEPFPFGEFLASGSSDANLKIWDIRKKG 143
QY 440 IVQLLOGEHLRCWPPHRTLRGRNKVTCCLLYPHQVSARYDORYLISGVDPFSVIWDI 499
Db 144 CIQ-----TYKHSRGISTIRTP-----DGRWYVSGLDNVVKVWDL 181
QY 500 FSGEMKHIFCVHGEITQLLVPPENCARSVOHCICSVASDHSVGLLSLRKKCIMLASRH 559
Db 182 TAGLHLHFKEFHGPIRSDFHP-----LEFLATGASADRTVKFWDLETFELIGSTRPE 235
QY 560 LFPQIVIKWRPS-----DDYL-----VVGCS- 580
Db 236 ATGVRISIKHPDGTTLFCGLDLSLKYVSWPEVVCVDMGWSLTLGDLCTISEGKLACSV 295
QY 581 -DGSVVVWQMD-----TGALDR-CWNGITAVAILNACDE----- 612
Db 296 YQNSVGIVWSDISQIBPYGIGSADKKECV-----EKILSALDQSPDRKSTPRSSSPD 350
QY 613 -----AVPRAVDLSHPAVNLKQAMTFRSLAALKKNWAHKLQTLATN 654
Db 351 YETKEIKNIYIDSTGNSAVAHKSGSLSTPATSTGQGNKSLV-----HSVPRSDSD 404
QY 655 LLASEASDKGN-----LPKY-----SHNSLMVQAIKNLTPD--- 686
Db 405 -ICKSDSDSGKSITFTSKTPGMLRPAVYKRTPTKFDETKKQSVAGVLYKSGLDGEKK 463
QY 687 -DIHVLFP-----FDVEALIIQLLT---EASRPNT-----ALISPENLQKA--- 723

Db 464 LDTETAFDSEMSGRNPYDADDSIIKSIITNKFQALLPESPTDEAKCMLKPPRVQSPST 523
QY 724 -----SGSDKGGSPFLTGRAAVLFQOVKE-----TIKENIK-----EHL 758
Db 524 KYNEARWATSDSGALDSKKNGLSSRDMDLPTGLADRGSNFCBEDIENKSISSRSERV 583
QY 759 LDDEEDBEIMQRRESDP-----EYRSSKSKPIUTLE--YNLTMDTAKLPMSCILHAWGLN 813
Db 584 LSPKAGDELKKSRSRSGSNHFIIGYEFYVQALNPLEDMRHLVIDSLK----- 631
QY 814 EVLDEVCLDRIGMLKPHCTVSGLLSRGGHMSLMLPGYNQPACKLSHGKTEVGRKLPASE 873
Db 632 -----TAVFAKVVSGRTRSLVERFERGE-KITHSEDKA-----ASA 667
QY 874 GVGKGTGYGSRAVTTQHLLSIISLANTLSM-----TNATFIGDHMKKGTPRPRSTPD 928
Db 668 TVVHSSNSVEERPLTASVQTVSMMPQVMPVKLDQATNSTTV-----DVPV 713
QY 929 LS--KARGSPPTSSNIVQGIQVAA---PVVSARSADHSGSDPPSPAPALHTCFLVNEG 983
Db 714 LSTRRTKSTFVRVMPVVLGRDTSMATDTPPTVSTRDRTSATN----- 756
QY 984 WSQLAHMCVLPDLGLDKFR-----PPLLEMLARR-----WQDR--CL 1021
Db 757 -----LTSDSVSGVTSKQRTTSPAPVMPMLNQTTMKKSDPSITSTWPDRTSAT 806
QY 1022 EVREAAQALLAELRRIEQGRKEAIDAWAPYLPQVIDHVI-----SPGVTS-----EAA 1071
Db 807 DLTSVDSVSGVTSRQTSTP-----APVPMKLNQTKIKSDPEPITSTRDPSA 856
QY 1072 QITT--APDAGSPEAKV-----QEEHDLVDDD-----ITTCGLSSVPO- 1109
Db 857 TNLTSDESPVTSROAKTSPAPVTPVILNQRTTNKMSDEPPVISTRPLRTSSARVMPVI 916
QY 1110 MKKISTSYEER-----RKQATAVLGVGAEGABIEPKLITRPS-- 1154
Db 917 LNOASTYDERPLSSRSRARTSPARIMPKLNQADNMPSEYEPVALTRSARNPARVIVP 976
QY 1155 QIPEGGGLTSGGSN-----YSLARHTCKALTF-----LLOPPSP 1189
Db 977 KLNQATNVTADASHIRSRORFPTQTLATPAVFDQVDMTLDDETTKQSSDILTQKEEP 1036
QY 1190 KLPPHTIRRTAIDLGR--GFTWPEYMDVSAVLMGELLELACADAEKQLANITMGLPLSP 1247
Db 1037 QIS-----GREDDGDIMEILMRTHSEVLNTL-----QSRLTKQLVAVFGT 1077
QY 1248 ADSARSARHALSLATARPAPFITTIAKEVH-RHTALAANTQSQNMHTTLAR----- 1301
Db 1078 VIQSTVSARVVGVLDLHAEERYLYFRLLLYLHVTHISLFRESKTNRELQVTLPLFCGI 1137
QY 1302 AKG 1304
Db 1138 AKG 1140

RESULT 15
S59317
DIP2 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein L3116; protein YLR129w
C:Species: Saccharomyces cerevisiae
C:Date: 29-Nov-1995 #sequence_revision 23-Feb-1996 #text_change 16-Aug-2004
C:Accession: S59317; S64971; S64966; S69416
R:Delius, H.
submitted to the EMBL Data Library, June 1995
A:Description: 36.8 kb of S.cerevisiae chromosome XII including ACE2, CK11, PDC5, SLS1
A:Reference number: S59317
A:Accession: S59317
A:Molecule type: DNA
A:Residues: 1-943
A:Cross-references: UNIPROT:Q12220; EMBL:X91258; NID:g995686; PIDN:CAA62640.1; PID:g995
R:Delius, H.; Hebling, U.

submitted to the Protein Sequence Database, May 1996

A;Reference number: S64967
A;Accession: S64971
A;Molecule type: DNA
A;Residues: 1-943 <DEW>
A;Cross-references: EMBL:Z73301; NID:g1360538; PIDN:CAA97699.1; PID:e245568; PID:g136053
A;Experimental source: strain S288C
R;Verhasselt, P.; Voet, M.; Volckaert, G.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64943
A;Accession: S64966
A;Molecule type: DNA
A;Residues: 1-458 <VER>
A;Cross-references: EMBL:Z73301; MIPS:YLR129w
A;Experimental source: strain S288C
R;Verhasselt, P.; Volckaert, G.
submitted to the EMBL Data Library, September 1995
A;Reference number: S69393
A;Accession: S69416
A;Molecule type: DNA
A;Residues: 1-458 <VEW>
A;Cross-references: EMBL:X89514; NID:g1297019; PIDN:CAA61707.1; PID:e198829; PID:g129704
C;Genetics:
A;Gene: SGD:DIP2
A;Cross-references: SGD:S0004119; MIPS:YLR129w
A;Map position: 12R
C;Superfamily: WD repeat homology
F;117-150/Domain: WD repeat homology <WD1>
F;200-231/Domain: WD repeat homology <WD2>
F;387-419/Domain: WD repeat homology <WD3>
F;569-602/Domain: WD repeat homology <WD4>

Query Match 2.3%; Score 182; DB 2; Length 943;

Best Local Similarity 20.2%; Pred. No. 0.00037;

Matches 139; Conservative 96; Mismatches 239; Indels 214; Gaps 33;

QY	10	IVLWGRKAPT-----HCISAVLLTDD--GATIVTGHCHDQICLDLSVLEQLINPRALL	60
Db	103	IKVWDLMSKTVLNFNGHKRAITLLQFDGTGTRLSIGSKDSNIIVWDLVGEVGLYK---L	159
QY	61	FGHTASITCLSKACASDKQYIVVSASESGEMCLWDVSDGRCIEFTKLACTHTG-----	113
Db	160	RSHKDSIT--GFWCQGED--WLSTSKDGMIKLWDLKTHQICB-THIA--HTGECWGLAV	212
QY	114	-----IQYQFSVGNQREGRLCHGHYPEILLVVDATS-----LEVLYSLVS	154
Db	213	KDOLLITTTGDSQVKIWKLDIENDKMG-----GKLTGEMGIFEKQSKQKGLKIEFITNSSD	267
QY	155	KISPDWISSM-STIRSHRTQEDTFVALSVTGILKMWITVSEISDMQDTPPIPEESKPIY	213
Db	268	KTSFFYIQNADKTIEFRIRKEEIA---RGLKK---REKRLKEKGLTE---EIAKSI-	317
QY	214	CONCQSIQCAF-----TORSLLVCSKYVRVFDAGDYSLLCSPSENGQTWTGGDFVS	267
Db	318	-----KESYSSFILHPFQTIRSLRYKIKSASWTT-----VS	347
QY	268	SDK--VLIITENGQSIYKL-----PASCPLAS-----DSFRSDVGKAVENLI	313
Db	348	SSKLELVLTSSNTIETIYISIPYKRDPTSPALKTHITIELQQTGTDVRS-----	396
QY	314	ILLDRDKKELLICPPVTRFPYCGREYFHKLLIQGDSGRINIWNISDTADKQSGSEGLAM	373
Db	397	--IDISDD-----NKLATA-SNGSLKIWNKTHKICIRTFECGYAL	434
QY	374	TTISISLQEAQDKLNPCPAGII-----DQLSVIPNSNEPLKVTA-----SVYIPA	417
Db	435	TKCF-----LPGGLLVILGTRNGELQFLDASSSLDTIEDHDAAIWSLDLTS	483
QY	418	HG-RLVCGREDGSI-----VIVPATQTAIVQLLQGEH-----MLRGRWPPH	457
Db	484	DGKRLVTGSADKTVKFWDFKVENSLVPGTKNFKLPVLKHHDTTLELTDDILCVRVSPDD	543
QY	458	R-----TLRGRHNKVTCL-----LYPHQ-----VSARYDQRYLISGGVDFSVIWDIF	500

Db	544	RYLAISLLDNTVKVFFLDMSKFPVLSLYGHKLPVLSIDISFDSKMIITSSADKNIKIWGLD	603
QY	501	SGEMKHIFCVHSGEITQLLVPPENCARVQHCHICSVASDHSVGLLSLRKKKCIIMLASRHL	560
Db	604	FGDCHKSLPAHQDSIMNVKFLPQS-----HNFSCSDAVVVKYWDGKFEKCIQKLYAHQ	657
QY	561	PTQVIKWRPSSDDYLVVVGSDGSGSVVWQ	588
Db	658	SEVWALAVATDGGFVVSSSHDSIRIWE	685

Search completed: June 30, 2005, 17:24:27

Job time : 62 secs

Db 312 VS-----GSDRTVKVMSTVDGSLHTLOQHTSTVRCMAGSILVTGSRDRTTLRVMDVE 366
Qy 445 QGEHMLRRGWPHRTILGRNKVTCLLYPHQVSARDQRLISGVDFSVIINDISGEM 504
Db 367 SGRHL-----ATLGHHAARVCV-----QPDGTTVSGGYDFTVKIWNHTGRC 410
Qy 505 KHIFCVHGGBITOLLVPPEN---CSARV-----QHCIC----- 534
Db 411 IRLTGHNRVYSLPESERSIVCSGLDTSIRVDFTRPEGQECVALLQCHTSLTSGMQ 470
Qy 535 -----SVADSHVGLLSREKKICMLASRHLFPIQVIKWRPDDYLVVGCSDGSGVYVW 587
Db 471 LRGNILVSCNADSHVRVMDIHEGTCVHMLSGHSAITSLQWP-GRNVATSSDDGTVKLW 529
Qy 588 QMDTGALDR-----CWGITAVEILNAC 610
Db 530 DIERGALIRDLVTLDSGGNGGCIWRLCSTSTMLAC 564

RESULT 2

US-08-751-189-4
; Sequence 4, Application US/08751189
; Patent No. 5919656
; GENERAL INFORMATION:
; APPLICANT: Harrington, Lea A.
; APPLICANT: Robinson, Murray O.
; TITLE OF INVENTION: No. 5919656el Genes Encoding Telomerase Protein
; TITLE OF INVENTION: 1
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen, Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,189
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleksi, Nancy A.
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-433
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2629 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-751-189-4

Query Match 2.1%; Score 166; DB 2; Length 2629;
Best Local Similarity 21.2%; Pred. No. 4.7e-05;
Matches 143; Conservative 88; Mismatches 235; Indels 210; Gaps 40;
Qy 12 LMRKAPTHCISAVLLTDDGATVTCGCHQICLWLSVLEQLNPRALLFGHTAS----- 66
Db 2065 LRHGEGVCCS---FSPDGGILATAGRDNLCLWDMKI-----AQAPLLIHTFSSCHRD 2116
Qy 67 -ITCLSKACASSDKQYIVSASBEGMCLWDVSDGRCIEFTKLAHTHTGIQFYQFS----- 120
Db 2117 WIT-----GCATWKDNLVSCSDGSGVCLWNPEAGQ-----QLGQFSGHQSA 2158
Qy 121 -----VGNQREGRLCHGHVPEILVVDATSLVLYSVLSKIS-----PWISS 163
Db 2159 VSAVVAVEEHIVSVSRDGLTKVWDHQ-----VELTSTPAHSGPISQCAAALEPRGGQPG 2214

Qy 164 MSIIRSHRTQEDTVVALSVTGILKVM--IVTSEISDMQD-----TPIFEESKPIYCON 216
Db 2215 SEL-----VTVGLDGTATKLWHPLLVQCIQRTLOHSGSPVTAASASGLLLTSD 2265
Qy 217 COSIS9FC-----AFTORSLLVVCSSKYRVRFDAGDYSLLCSPSENGQ--TWGQDFV 266
Db 2266 DSSVQLWQIPKEADDSYKPRSSVAITAVAW---APDGSMMVVG-NEAGELTLWQQAQAV 2320
Qy 267 SSDKV-----IIV-----TENQSQSYIYKLPASCLPASDS-----FRSDVGKA 303
Db 2321 ATAQAPGRVSHLIWYSANSFFVLSEANENVSEWQVGLRKGSTSTSSSLHLKRLVLOEDWGL 2380
Qy 304 VE-NLIPPVQHTLLDOKKELLICPVTTRFFYGCREY-FHKLLI-----QGPS 349
Db 2381 TGLGLAPDQGSLLMKEDVELLEMKPGSIPSSICRRYGVHSSILCTSKYGLFYLOQGD- 2439
Qy 350 SGRLLNIWNISDTADKQSGSEGLAMTTSISLOBAFQKLNPCPAGIIDLQLSVIFNSNPLKV 409
Db 2440 SGLLSILEQESGE---PEEILDENLNLN-----NP-----NGSPVSI 2474
Qy 410 TASVYIPAHGRVLCGREDGSIIVIVPATQTAIVQLQGEHMLRRGW-----PPHRTLGRH 463
Db 2475 T-QAKPESESSLICATSDGMLWNLSCTS-----EGEWIVDNIWQKAKKPKTOTLETE 2527
Qy 464 RNKVTCLLYPHQVSARYDQRYLISGVDPFSVIINDIFSGEMKHIFC--VHGGEITOLLVP 521
Db 2528 -----LSPH-----SELDFSIDCW-IDPTNLKAOQCKKHLHGSVTAHVL 2566
Qy 522 PENCARSVQHCICSVADSHVGLLSREKKICMLASRHLF-----PIQVIK-W-RPSDD-Y 574
Db 2567 P-----GLLVTASKORDV---KLWERPSQLLGG--LFRCEGVPVSCLEPNWEPSPPLQ 2613
Qy 575 LVVGCSDGSGVYV--WQ 588
Db 2614 LAVGDTQGNLYFLSWE 2629

RESULT 3

US-09-060-836-4
; Sequence 4, Application US/09060836
; Patent No. 5981707
; GENERAL INFORMATION:
; APPLICANT: Harrington, Lea A.
; APPLICANT: Robinson, Murray O.
; TITLE OF INVENTION: No. 5981707el Genes Encoding Telomerase Protein
; TITLE OF INVENTION: 1
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen, Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/060,836
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/751,189
; FILING DATE: 15-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleksi, Nancy A.
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-433
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:

LENGTH: 2629 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-09-060-836-4

Query Match 2.1%; Score 166; DB 2; Length 2629;

Best Local Similarity 21.2%; Pred. No. 4.7e-05;
Matches 143; Conservative 88; Mismatches 235; Indels 210; Gaps 40;

QY 12 LWRKAPTHCISAVLLTDDGATVTCGHDGQICLDLWLSVELQINPRALLFGHTAS-----66
DB 2065 LRHGEFVCCS---FSPDGGILATAGDRNLLCWDNKI-----AQAPLLIHTFSSCHRD 2116
QY 67 -ITCLSKACASSDKQYIVSASEGEMCLWDVSDGRCIEFTKLAETHGTFYQFYS-----120
DB 2117 WIT---GCAWTKDNILVSCSDSGVGLWNPAGQ-----QLGQFSGHQA 2158
QY 121 -----VGNQREGRLCHGHYPEILVVDATSLVLYSLVSKIS-----PDMISS 163
DB 2159 VSAVVAVEEHIVSVSRDGLTKVMDHQG---VELTSIPAHSGPISQCAAALEPRPGQPG 2214
QY 164 MSIIRSHRTQEDTVAVALSVTGLKW---IVTSISDMQD-----TEPIPEESKPIYQCN 216
DB 2215 SELL-----VVTVGLDGATKLHPLLVCOIRTLQGHSGPVTAAASEASGLLITSD 2265
QY 217 CQGISFC-----AFTORSLLVVCVKYRVPDAGDYSLLCSGSPSNGQ--TWTCGDFV 266
DB 2266 DSSVOLWQIPKEADDSYKPRSSVAITAVAW---APDGSVMVSG-NEAGELTLWQQA 2320
QY 267 SSDKV-----IIW-----TENGQSYIYKLPASCLPASDS-----FRSDVGKA 303
DB 2321 ATAQAPGRVSHLIWYSANSPFVLSANENVSQVGLRKGSTSTSSSLHLKRVLQEDWGV 2380
QY 304 VE-NLIPVQHILLDRKDKELLICPPVTRFFYGCY-FAKLLI-----QGDS 349
DB 2381 TGLGLAPDQGSLLMKEDVELLEMKPGSIPSSICRRYGVHSSILCTSKYGLFYQGD- 2439
QY 350 SGRINLWNIISDTADKQSGSEGLMTTTSISLQEAFLKPNCPAGIIDQLSVIPNSNPLKV 409
DB 2440 SGLLSILEQESGE---FEIILDFNLN-----NP-----NGSPVSI 2474
QY 410 TASVYIPAHGRVLCGRDGSIVVPATQTAIVOLLQGEHMLRGW-----PPHRTLGRH 463
DB 2475 T-QAKPESSESLLCATSDGMLNLSCTS-----EGEIVDNIWQKAKKPKTQLETE 2527
QY 464 RNKVTCLLYPHQVSARDYRLISGGVDFSVIWDIFSGEMKHIFC--VHGGEITQLLVP 521
DB 2528 -----LSPH-----SELDFSIDCW-IDPTNLKQAQCKKHLSGVTALHVL 2566
QY 522 PENCARSVQHCISVASHSVGLLSLREKKICMLASHLF---PIQVIK-W-RPSDD-Y 574
DB 2567 P-----GLLVTASKDRV---KLWERPQMQLLG--LPRCGPVSCLPEWNEPSPQLQ 2613
QY 575 LVVGCSDGSVYV--WQ 588
DB 2614 LAVGDTQGNLYFLSWE 2629

RESULT 4

US-09-184-445-4
Sequence 4, Application US/09184445
Patent No. 6174703
GENERAL INFORMATION:
APPLICANT: Harrington, Lea A.
APPLICANT: Robinson, Murray O.
TITLE OF INVENTION: No. 6174703el Genes Encoding Telomerase Protein
TITLE OF INVENTION: 1
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen, Inc.
STREET: 1840 De Havilland Drive

CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/184,445
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/751,189
FILING DATE: 15-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Oleski, Nancy A.
REGISTRATION NUMBER: 34,688
REFERENCE/DOCKET NUMBER: A-433
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2629 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-09-184-445-4

Query Match 2.1%; Score 166; DB 3; Length 2629;

Best Local Similarity 21.2%; Pred. No. 4.7e-05;
Matches 143; Conservative 88; Mismatches 235; Indels 210; Gaps 40;

QY 12 LWRKAPTHCISAVLLTDDGATVTCGHDGQICLDLWLSVELQINPRALLFGHTAS-----66
DB 2065 LRHGEFVCCS---FSPDGGILATAGDRNLLCWDNKI-----AQAPLLIHTFSSCHRD 2116
QY 67 -ITCLSKACASSDKQYIVSASEGEMCLWDVSDGRCIEFTKLAETHGTFYQFYS-----120
DB 2117 WIT---GCAWTKDNILVSCSDSGVGLWNPAGQ-----QLGQFSGHQA 2158
QY 121 -----VGNQREGRLCHGHYPEILVVDATSLVLYSLVSKIS-----PDMISS 163
DB 2159 VSAVVAVEEHIVSVSRDGLTKVMDHQG---VELTSIPAHSGPISQCAAALEPRPGQPG 2214
QY 164 MSIIRSHRTQEDTVAVALSVTGLKW---IVTSISDMQD-----TEPIPEESKPIYQCN 216
DB 2215 SELL-----VVTVGLDGATKLHPLLVCOIRTLQGHSGPVTAAASEASGLLITSD 2265
QY 217 CQGISFC-----AFTORSLLVVCVKYRVPDAGDYSLLCSGSPSNGQ--TWTCGDFV 266
DB 2266 DSSVOLWQIPKEADDSYKPRSSVAITAVAW---APDGSVMVSG-NEAGELTLWQQA 2320
QY 267 SSDKV-----IIW-----TENGQSYIYKLPASCLPASDS-----FRSDVGKA 303
DB 2321 ATAQAPGRVSHLIWYSANSPFVLSANENVSQVGLRKGSTSTSSSLHLKRVLQEDWGV 2380
QY 304 VE-NLIPVQHILLDRKDKELLICPPVTRFFYGCY-FAKLLI-----QGDS 349
DB 2381 TGLGLAPDQGSLLMKEDVELLEMKPGSIPSSICRRYGVHSSILCTSKYGLFYQGD- 2439
QY 350 SGRINLWNIISDTADKQSGSEGLMTTTSISLQEAFLKPNCPAGIIDQLSVIPNSNPLKV 409
DB 2440 SGLLSILEQESGE---FEIILDFNLN-----NP-----NGSPVSI 2474
QY 410 TASVYIPAHGRVLCGRDGSIVVPATQTAIVOLLQGEHMLRGW-----PPHRTLGRH 463
DB 2475 T-QAKPESSESLLCATSDGMLNLSCTS-----EGEIVDNIWQKAKKPKTQLETE 2527
QY 464 RNKVTCLLYPHQVSARDYRLISGGVDFSVIWDIFSGEMKHIFC--VHGGEITQLLVP 521
DB 2528 -----LSPH-----SELDFSIDCW-IDPTNLKQAQCKKHLSGVTALHVL 2566

```
QY 522 PENCARSVORHCISVASHDHSVGLLSLRKKCICMLASHLF-----PIQVIK-W-RPDSDD-Y 574
Db 2567 P-----GLVTASKDRV---KLWERSMQLLG---LPRCGPVSCLPWNPEPSPLQ 2613
QY 575 LVVGCSDGSVYV--WQ 588
Db 2614 LAVGDTQGNLYFLSWE 2629

RESULT 5
US-09-092-508-2
; Sequence 2, Application US/09092508
; Patent No. 6291643
; GENERAL INFORMATION:
; APPLICANT: Henzel, William J.
; TITLE OF INVENTION: APAF-1, AN ACTIVATOR OF C ASPASE-3
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 6291643west Center, 90 South Seventh St
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092,508
; FILING DATE: 05-JUN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/048,807
; FILING DATE: 05-JUN-1997
; APPLICATION NUMBER: 60/055,258
; FILING DATE: 07-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Kettelberger, Ph.D., Denise M
; REGISTRATION NUMBER: 33,924
; REFERENCE/DOCKET NUMBER: 11669.6USU1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1194 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-09-092-508-2

Query Match 2.1%; Score 163; DB 3; Length 1194;
Best Local Similarity 19.8%; Pred.No.2e-05;
Matches 198; Conservative 135; Mismatches 314; Indels 354; Gaps 56;

QY 115 QYQFQSVGNOREGRLCHGHPYELVVDATSLVLSVSKISFPDWTSSMSIIRSHRTOE 174
Db 328 EYILKQLNQKQFRIRKSSSYDEALDEANSISV--EMLREDIKDYTDLSIL-----QK 380
QY 175 DTVVALLSVTGILKWIVTSHDMDQTEPIFEESKPIYQNCQSIQSFCAFTQRSLLVVC 234
Db 381 DVKVPKVLKIL--WDMETE-----EVEDILQE-----FVNKSL-FC 415
QY 235 SKYWRVFDAGDYSLLCSGSPENGQWTGGDFVSDKVIWITENGQSVI-----282
Db 416 DR-----NGKSPR--YYLHDLQVDFLTERKNCQSQQLDKHKIITQFQ 454
QY 283 -YKLPAACLPAQSDS-----FRSDVGKAVENLIPPVQHI--314
```

```
Db 455 RYHQPHTLSPDQSDQCMYWTNFLAYHMASAKMHKELCALMFSLDWIKAKTELVCAPAHLE 514
QY 315 -----LLDRKKKELLI CPPVTRFFYGCREYFHKLLIQDSSGRLNIWI-----SDT 361
Db 515 FVEYRHILDEK-----CAVSENF-----QEF-----LSLNGHLIGRQFPFNIQGLCEPET 562
QY 362 AD-----KQGEGLAMTTTSLQAFKPLNCPAGIIDQLSVINSNEPLKVTAS 412
Db 563 SEVYQAKLOAKQEVNDGMLYLEWINKNITN-----LSRLVVRH-----TDA 606
QY 413 VYIPAHGRIVCGREDS-IVIVPATQT-AIVQLLOGEHMLRRGWPPHRTLGRHNKVTCL 470
Db 607 VY---H---ACPSDQRIASCADKTLQVFAETGEKLE-----IKAHEDEVLC 652
QY 471 LYPHQVSARYDORYLISGVSFVIWIDIFSEGMKHFVCHGGEITQLLVPPENC---SA 527
Db 653 AF-----STDDRFIATCSDVKVKIWNMSMTGELVHTYDEHSEQ-----NCCHFTN 698
QY 528 RVQHCICSV-ASDHSVGLLSLRKKCICMLASHRLLPPIQVIKWRPSDDYLWVGCs-DGSVY 585
Db 699 SSHLLLATGSSDDCFLLKMDLNQKCRNTMFGHTSVNHCRFSP-DDKLLASCSDGTLYK 757
QY 586 VMQMDTGALDRCVMGITAVEILNACD--EAVPAADVSLSHPAVNLKQAMTRSLAALKM 643
Db 758 LWDA-TSANERKSINVKQF-FLNLEDPQEDMEVIVKCCSWSADG-----ARIMVAAKN- 808
QY 644 AHKLOTLATNLLASEADKGNLPKYSHNSLWQAIKTNLTDPD-----687
Db 809 ---KIPLMNTDSRSKVADCRGHL-SWVHGVMFSPDGSSFLTSSDDQTIRLWETKVKCKNS 864
QY 688 -----IHVLFDFVEALI-----IQLLTEASRPNTALISPENLQK 722
Db 865 AVMLKQEVDDVFQENVEVMVLAVDHRRLQLINGRGQIDILYE--AQVSCCLSPHLOYI 922
QY 723 ASGSSDKGGSF-----LTGKRAAVLFQ---QVKET---IKENIKEHLDDDEEDEE---767
Db 923 AFG--DENGAIILELVNRR---IFQSRFOHKKTWVHIQFTADEKTLISSDDAEIQVWN 977
QY 768 -----IMQRRESDPEYRSKSKPLTLELVNLTMDTAKLPWSCILHANGL---NEVLDE 818
Db 978 WQLDKCIFLRGHQETVKQFRLKNS--RLLSWS-----FDGTVKVWNIIITGNKEKDF 1027
QY 819 VCLDRGLMKPHCTVSFGLLSRGHMSLMLPGVNOPACKLSHGKTEVGRK-----868
Db 1028 VC-----HQTVL-----SCDISHATKFSSTADTKAKIWS 1059
QY 869 ----LPASEGVGKTYGVSRVTTQHLJSIIISLANTLMSMTNATFIDHMKKGPTPRRP 924
Db 1060 FDLILLPLHE--LRGHNGCVRC-----SAFVSDSTLLA-----TGD-----1092
QY 925 STFDLSKARGSPPTSSNIVQGOIKQVAAAPVVSARSADADHSG 965
Db 1093 ---DNGEIR-----IWNVSNGLLHLHLCAP-LSEGAATHGG 1124

RESULT 6
US-09-435-115-2
; Sequence 2, Application US/09435115
; Patent No. 6346607
; GENERAL INFORMATION:
; APPLICANT: Henzel, William J.
; TITLE OF INVENTION: APAF-1, AN ACTIVATOR OF C ASPASE-3
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 6346607west Center, 90 South Seventh St
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```


COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/435.115
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/092,508
FILING DATE:
APPLICATION NUMBER: 60/055,258
FILING DATE: 07-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Kettelberger, Ph.D., Denise M
REGISTRATION NUMBER: 33,924
REFERENCE/DOCKET NUMBER: 11669.6USU1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1194 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-435-115-2

Query Match 2.1%; Score 163; DB 3; Length 1194;
Best Local Similarity 19.8%; Pred. No. 2e-05;
Matches 198; Conservative 135; Mismatches 314; Indels 354; Gaps 56;
QY 115 QYQFQSVGNREGRLCHGHPYELVVDATSLVLSVSKISPDWISSMSIIRSHRTOE 174
DB 328 EYLLKQLQNKQFKIRKSSDYDEALDEAMSISV--EMLRDIDKYVYDLSIL-----QK 380
QY 175 DTVALSVTGILKWIIVTSISDMQDTEPIFEESKPIYQNCQISFCFACFTORSLLVVC 234
DB 381 DVKVPVKVLCIL--WDMETE-----EVEDILQE-----FVNKSL-FC 415
QY 235 SKYWRVPDAGDYLSCGSPENGQWTGDFVSSDKVVIWTENGQSVI----- 282
DB 416 DR-----NGKSPR--YYLHDLQVDFLTKNCSQQLDLHKKIITQFQ 454
QY 283 -YKLPASCLPASDS-----FRSDVGKAVENLIPVQHI-- 314
DB 455 RYHQPHLSPDQEDCMYWNFLAYHMASAKMKELCALMFSLDWIKAKTELVGPAHLIHE 514
QY 315 -----LLDRKDKELLIICPPVTRFFYGCYREYFHKLLIQGDSGRLNIWNI-----SDT 361
DB 515 FVEYRHLDEKD-----CAVSENF-----QEF--LSLNGHLGROPPNIVQLGCEPET 562
QY 362 AD-----KQSGEGLAWTTSISQEAFLKINPCPAGIIDQLSVIPNSNPLKVTAS 412
DB 563 SEVYQAKLQAKQEDVNGMLYLEWINKNITN-----LSRLVVRPH-----TDA 606
QY 413 VYIPAHGRLVCGREDGS-IVIVPATQF-AIVQLQGEHMLRRGPPHRTLRGHNKVTCL 470
DB 607 VY---H---ACFSESDGRIASCGADKTLQVFKATPGEKILLE-----IKAHEDEVLC 652
QY 471 LYPHQVSARYDQRLYISGGVDFSVIWIWDFSGEMKHIFCVHGGEITQLLVPPENC---SA 527
DB 653 AP-----STDDRFIATCSVDKVKIWNMTGELVHYDSEQV-----NCCHFTN 698
QY 528 RVQHCICSV-ASHSVGLLSLRKKCIIMLASRHLFPIQVIKWRPSDDYLVGCS-DGSVY 585
DB 699 SSHLLLATGSSDCFLKLDLWLNQKCRNTWFGHTNSVNHCRFSP-DDKLLASCASDGTLK 757
QY 586 VQWMDTALDRCVNGITAVELIACD--EAVPAVDSLSHPAVNLQAMTRRSALAKNM 643
DB 758 LWDA-TSANERKGINVKQF-FLNLEDQEDMEVIVKCCSWSDG-----ARIMVAKN- 808

QY 644 AHKKLOTATNLLASEASDKGNLPKYSHNSLAWQAIKTNLTDPD----- 687
DB 809 ---KIFLWNTDSRSKVADCRGHL-SWVHGVMFSPDGSSFLTSSDDQITRLWETKCKVCKNS 864
QY 688 -----IHVLPFDVEALI-----IQLLTEASRPNTALISPENLQK 722
DB 865 AVMLKQEVDFVQFENEVMVLAVDHIRRLQLINGRTGQIDYLTE--AQVSCCCLSPHLOYI 922
QY 723 ASGSSDKGGSF-----LTGKRAAVLFQ---OVKET---IKENIKHEHLIDDEEEDDEE----- 767
DB 923 AFG--DENGAIETLELVNRR---IFQSRFOHKKTVMHIOFADKTKLSSDDDAEIQWVN 977
QY 768 -----IMQRRESDPEYRSSKPLTLLVNLTMWTAKLFMSCLHAWGL---NEVLDE 818
DB 978 WQLDKICIFLRGHQETVKDFLLKNS--RLLSWS-----FDGTVKVMNIITGNKEKDF 1027
QY 819 VCLDRGLMKPHCTVSVFGLLSRGHMSLMLPGYNQPACKLSHGKTEVGRK----- 868
DB 1028 VC-----HQTGVL-----SCDISHDATKFSSTSAKTAKIWS 1059
QY 869 ---LPASEGVGKGTYGVSRAVTTQHLLSIIISLANTILMSMTNATFIGDHMKKGPTRRPRP 924
DB 1060 FDLILLPLHE--LRGHNGCYRC-----SAFSDVSTILLA-----TGD----- 1092
QY 925 STPDLSKARGSPPTSSNIYVQGIQKQVAAPVVSARSADHSG 965
DB 1093 ---DNGBEIR-----IWNVSNGELLHLCAP--LSEGAATHGG 1124
RESULT 7
US-09-069-023-26
; Sequence 26, Application US/09069023A
; Patent No. 6348573
; GENERAL INFORMATION:
; APPLICANT: Nunez, Gabriel
; APPLICANT: Inohara, Naohiro
; APPLICANT: Koseki, Takeyoshi
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS
; TITLE OF INVENTION: SIGNALING PATHWAY INHIBITORS AND ACTIVATORS
; FILE REFERENCE: UM-0333
; CURRENT APPLICATION NUMBER: US/09/069,023A
; CURRENT FILING DATE: 1998-04-27
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 26
; LENGTH: 1194
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-069-023-26

Query Match 2.1%; Score 163; DB 3; Length 1194;
Best Local Similarity 19.8%; Pred. No. 2e-05;
Matches 198; Conservative 135; Mismatches 314; Indels 354; Gaps 56;
QY 115 QYQFQSVGNREGRLCHGHPYELVVDATSLVLSVSKISPDWISSMSIIRSHRTOE 174
DB 328 EYLLKQLQNKQFKIRKSSDYDEALDEAMSISV--EMLRDIDKYVYDLSIL-----QK 380
QY 175 DTVALSVTGILKWIIVTSISDMQDTEPIFEESKPIYQNCQISFCFACFTORSLLVVC 234
DB 381 DVKVPVKVLCIL--WDMETE-----EVEDILQE-----FVNKSL-FC 415
QY 235 SKYWRVPDAGDYLSCGSPENGQWTGDFVSSDKVVIWTENGQSVI----- 282
DB 416 DR-----NGKSPR--YYLHDLQVDFLTKNCSQQLDLHKKIITQFQ 454
QY 283 -YKLPASCLPASDS-----FRSDVGKAVENLIPVQHI-- 314
DB 455 RYHQPHLSPDQEDCMYWNFLAYHMASAKMKELCALMFSLDWIKAKTELVGPAHLIHE 514
QY 315 -----LLDRKDKELLIICPPVTRFFYGCYREYFHKLLIQGDSGRLNIWNI-----SDT 361

RESULT 12

US-09-949-016-10066
; Sequence 10066, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10066
; LENGTH: 1196
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10066

Query Match 2.1%; Score 163; DB 4; Length 1196;

Best Local Similarity 19.8%; Pred. No. 2e-05;
Matches 198; Conservative 135; Mismatches 314; Indels 354; Gaps 56;

QY 115 QYFQSVGNQREGRILCHGYPIILVVDATSLVLSVSKISPDWISMSIIRSHRTQE 174
DB 330 EYLLQONKQFKRIKSSDYDEALDEAMSISV--EMLREDIKYITDLSIL-----OK 382
QY 175 DTVALSVTGILKVMIVTSEISDMQDEPIFEESKPIYCONQCSISFCACFORSLLVC 234
DB 383 DVKPTKVLICL--WMETE-----EVEDILQE-----FVNSLL-FC 417
QY 235 SKYWRVFDAGDYSLLCGSPSENGQTWTGDFVSSDKVVIWTENGQSYI----- 282
DB 418 DR-----NGKSPR--YLLHDLQVDFTEKNCQSLQDLHKKIITQFQ 456
QY 283 -YKLPAACLPASDS-----FRSDVGKAVENLPPVQHI-- 314
DB 457 RYQHPHTLSPQDCMYWYNFLAYHMASAKMKELCALMFSLDWIKAKTELVPALHIE 516
QY 315 -----LLDRKOKELICPPVTFFYGCYEFHKLIIQDSSGRINWNI-----SDT 361
DB 517 FVEYRHILDEK-----CAVSENF-----QEF--LSINGHLGRQPPNIVQLGCEPET 564
QY 362 AD-----KQGSBEGLAMTTSISLQAFADKLNPCPAGIIDQLSVIPNSNEPLKVITAS 412
DB 565 SEVYQAKLQAKQEVNDGMLYLEWINKNITN-----LSRLVVRPH-----TDA 608
QY 413 VYIPANGRLVCGREDGS-IVIVPATOT-AIVQLLQHEHMLRGPPIHRLGRHNKVTC 470
DB 609 VY-----H---ACFSEDQRIASCCGADTLQVFAETGEKLE-----IKAHEDEVLC 654
QY 471 LYPHQVSARVDORYLISGGVDFSVIITWDFSGEMKHFVHGGEITQLLVPPENC---SA 527
DB 655 AF-----STDDRFATCSQKVKIWNMTGELVHTYDEHSEV-----NCCHFTN 700
QY 528 RVQHICISV-ASPHSVGLLSLREKKICMLASRLHFPPIQVIKWRPDDYLIVVCGS-DGSVY 585
DB 701 SSHLLLATGSSDCFLKLDNLNKECRNTMFGHTNSVNHCRFSP-DDKLLASCADGTLK 759
QY 586 VQMDTGALDRCVMGITAVEILNACD--EAPVAAVDSLHPANLKOAMTRSLAALKNN 643
DB 760 LMDA--TSANERKSINKQF--FLNLEDPQEDMEVIVKCCSWSDG-----ARIMVAAKN- 810
QY 644 AHKLOTLATNLASEASDKGNLPKYSNLSLWQAIAKTNLTDPD----- 587
DB 811 ---KIFLWNTDSKRVADCRGHL-SWVHGYMFSPDGSSFLTSSDDQTIIRLWETKVKCKNS 866

QY 688 -----IHVLPFDVEALI-----IOLLTEBASRPNALISPENLOK 722
DB 867 AVMLKQEVDFQENEVMVLAVDHIRRLQILINGRTGQIDYLTE--AQVSCCLSPHLQYI 924
QY 723 ASSSDKGGSF---LTGKRAAVLFO---QVKET---IKENIKHLLDDEEED--- 767
DB 925 AFG--DENGAIILELVNRR---IFQSRFQHKKTVMHIOFTADEKTLISSDDAEIQVWN 979
QY 768 -----IMQRRESDPEYRSKSKPLTLLEYNLMTDTAKLFMSCLHAWGL---NEVLDE 818
DB 980 WQLDKCIPLRGHQETVKDFRLKNS--RLSNS-----PDGTVKVNNIITGNKEKDF 1029
QY 819 VCLDRGLMKLPKCTVTSFGLLSRGHMSMLPGVNPQACKLSHGKTEVGRK----- 868
DB 1030 VC-----HOGTVL-----SCDISHDATKFSSTSDAKTAKIWS 1061
QY 869 -----LPASEGVGKGTGVSRAVTTQHLLSIIISLANTLMSMTNATFIGDHMKGPTPRPP 924
DB 1062 FLLLLPLHE--LRGHNGCVRC-----SAFVSDSTLLA-----TGD----- 1094
QY 925 STPDLSKARGSPPTSSNIVQGOIKQVAAFPVVSARSADHSG 965
DB 1095 ---DNGEIR-----IWNVSGELLHLCAP-LSEEGAATHGG 1126

RESULT 13

US-09-092-508-16
; Sequence 16, Application US/09092508
; Patent No. 6291643
; GENERAL INFORMATION:
; APPLICANT: Henzel, William J.
; TITLE OF INVENTION: APAP-1, AN ACTIVATOR OF C ASPASE-3
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 6291643west Center, 90 South Seventh St
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092,508
; FILING DATE: 05-JUN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/048,807
; FILING DATE: 05-JUN-1997
; APPLICATION NUMBER: 60/055,258
; FILING DATE: 07-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Kettelberger, Ph.D., Denise M
; REGISTRATION NUMBER: 33,924
; REFERENCE/DOCKET NUMBER: 11669.6USU1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-092-508-16

Query Match 2.1%; Score 163; DB 3; Length 1205;
Best Local Similarity 19.8%; Pred. No. 2e-05;
Matches 198; Conservative 135; Mismatches 314; Indels 354; Gaps 56;

QY 115 QYQFSVGNQREGRLCHGHPBILVVDATSEVLYSLVSKISPDWISSMSIIRSHRTQE 174
DB 339 EYLLQQLQKQPKRIRKSSSYDYALDEAMSISV--EMLREDIKOYYTDLISL-----QK 391

QY 175 DTVVALSVTGILKVIWTVSEISDMODTEPIFEESKPIYQCONQSISFCQAFQSRLLVVC 234
DB 392 DVKVPVKVLCIL--WDMETE-----EVEDILOE-----FVNKSLL-FC 426

QY 235 SKYWRVFDAGDYSLLCSGSPENGOTWTGGDFVSSDKVITWTENGOSYI----- 282
DB 427 DR-----NGKSPR--YYLHDLQVDFLTEKNCSQLDLHKKIITQFQ 465

QY 283 -YKLPAASCLPASDS-----FRSDVGKAVENLIPPVQHI-- 314
DB 466 RYHQPHTLSPQEDCMYWNFLAYHMASAKMHKELCALMFSLDWIKAKTELGVPAHLIHE 525

QY 315 -----LLDRKDKELLICPPVTRFFYGCREFYFKLLIQDSSGRNLNWI-----SDT 361
DB 526 FVEYRHILDEKO-----CAVSENF-----QEF-----LSLNGHLLGRQPPFNIVQLGCEPET 573

QY 362 AD-----KQSBEGLAMTTSISLQBAFDKLNPCPAGIIDQLSVIPNSNEPLKVTAS 412
DB 574 SEVYQQAQKQAEVNGMLYLEWINKNITN-----LSRLVVRH-----TDA 617

QY 413 VYIAPAGRLVCGREDGS-IVIVPATQT-AIVQLLOGEHLRGRWPPHRTLRGHRNKVTL 470
DB 618 VY---H--ACFSESDGORIASCGADKTLQVFAETGEKLE-----IKAHEDEVLC 663

QY 471 LYPHQVSARYDQRYLISGGVDFSVIWDIFSGEMKHI FCHVGGEITQLLVPPENC---SA 527
DB 664 AF-----STDDRIFATCSVDKVKIWSMTGELVHTYDEHSEV-----NCCHFTN 709

QY 528 RVQHCICSV-ASDHSVGLLSLRKCKIMLASRLFPQIKWRPDDLYLVVCGS-DGSVY 585
DB 710 SSHLLLATGSSDCFLKLDLQKECNTWTFGHTNSVNHCRFSP-DDKLLASCADGTLK 768

QY 586 VQWQMDGALDRCVMGITAVEILNACD--EAVPAADVLSLHPAVNLKQAMTRSLAALKNM 643
DB 769 LWDA-TSANERKSINVQKQ-FLNLEDPQEDMEVIVKCCSWSADG-----ARIMVAAKN- 819

QY 644 AHKLQTLATNLASEASDKGNPKYSHNSLMVQAIKMTLTDPP----- 687
DB 820 ---KIFLWNTDSRSKVADCRGHL-SWVHGVMFSPDGSSFLTSSDDQTIRLWETKVCYKNS 875

QY 688 -----IHVLPFDVREALI-----IQLLTEASRPNTALISPENLQK 722
DB 876 AVMLKQEVDFVQENVWVLAVDHIRRLQINGRTQIDYLTE--AQVSCCCLSPHLQYI 933

QY 723 ASGSDKGGSF-----LTGKRAAVLFQ--QVKET-----IKENIKEHLDDDEEDEE----- 767
DB 934 AFG--DENGAIETELVNNR---IFQSRFQHKTVWHIQTADAETKLISSSDDAEIQVWN 988

QY 768 -----IMQRRESDEPERSKSLPLTLLYNLTMDTAKLFMSCLHAWGL-----NEVLDE 818
DB 989 WQLDKCIPLRGHOETVKDFRLLKNS--RLLSWS-----FDGTVKVWNITGNKEKDF 1038

QY 819 VCLDRGLMKLPHCTVSGLLSRGHSMLPGYNQACKLSHGKTEVGRK----- 868
DB 1039 VC-----HOGTVL-----SCDISHDATFESSTADKTAIWS 1070

QY 869 -----LPASEGVKGTVGSRAVTTQHLLSIISLANTLMGMTNATFIDGHMKKGTPRPPR 924
DB 1071 FDLLLPLHE--LRGHNGVCAC-----SAFVSDSTLLA-----TGD----- 1103

QY 925 STPDLSKARSGPTSSNIVGOIQKQVNAAPVVSARSADHSG 965
DB 1104 ---DNGEIR-----INWVNSGELLHLCAP-LSBEGAATHGG 1135

RESULT 14
US-09-435-115-16
; Sequence 16, Application US/09435115
; Patent No. 6346607
; GENERAL INFORMATION:
; APPLICANT: Henzel, William J.
; TITLE OF INVENTION: APAP-1, AN ACTIVATOR OF C ASPASE-3
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 6346607west Center, 90 South Seventh St
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/435,115
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/092,508
; FILING DATE:
; APPLICATION NUMBER: 60/055,258
; FILING DATE: 07-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Kettelberger, Ph.D., Denise M
; REGISTRATION NUMBER: 33,924
; REFERENCE/DOCKET NUMBER: 11669.6USU1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-09-435-115-16

Query Match 2.1%; Score 163; DB 3; Length 1205;
Best Local Similarity 19.8%; Pred. No. 2e-05;
Matches 198; Conservative 135; Mismatches 314; Indels 354; Gaps 56;

QY 115 QYQFSVGNQREGRLCHGHPBILVVDATSEVLYSLVSKISPDWISSMSIIRSHRTQE 174
DB 339 EYLLQQLQKQPKRIRKSSSYDYALDEAMSISV--EMLREDIKOYYTDLISL-----QK 391

QY 175 DTVVALSVTGILKVIWTVSEISDMODTEPIFEESKPIYQCONQSISFCQAFQSRLLVVC 234
DB 392 DVKVPVKVLCIL--WDMETE-----EVEDILOE-----FVNKSLL-FC 426

QY 235 SKYWRVFDAGDYSLLCSGSPENGOTWTGGDFVSSDKVITWTENGOSYI----- 282
DB 427 DR-----NGKSPR--YYLHDLQVDFLTEKNCSQLDLHKKIITQFQ 465

QY 283 -YKLPAASCLPASDS-----FRSDVGKAVENLIPPVQHI-- 314
DB 466 RYHQPHTLSPQEDCMYWNFLAYHMASAKMHKELCALMFSLDWIKAKTELGVPAHLIHE 525

QY 315 -----LLDRKDKELLICPPVTRFFYGCREFYFKLLIQDSSGRNLNWI-----SDT 361
DB 526 FVEYRHILDEKO-----CAVSENF-----QEF-----LSLNGHLLGRQPPFNIVQLGCEPET 573

QY 362 AD-----KQSBEGLAMTTSISLQBAFDKLNPCPAGIIDQLSVIPNSNEPLKVTAS 412
DB 574 SEVYQQAQKQAEVNGMLYLEWINKNITN-----LSRLVVRH-----TDA 617

QY 413 VYIAPAGRLVCGREDGS-IVIVPATQT-AIVQLLOGEHLRGRWPPHRTLRGHRNKVTL 470
DB 618 VY---H--ACFSESDGORIASCGADKTLQVFAETGEKLE-----IKAHEDEVLC 663

QY 471 LYPHQVSARYDQRYLISGGVDFSVIWDIFSGEMKHI FCHVGGEITQLLVPPENC---SA 527
DB 664 AF-----STDDRIFATCSVDKVKIWSMTGELVHTYDEHSEV-----NCCHFTN 709

QY 528 RVQHCICSV-ASDHSVGLLSLRKCKIMLASRLFPQIKWRPDDLYLVVCGS-DGSVY 585
DB 710 SSHLLLATGSSDCFLKLDLQKECNTWTFGHTNSVNHCRFSP-DDKLLASCADGTLK 768

QY 586 VQWQMDGALDRCVMGITAVEILNACD--EAVPAADVLSLHPAVNLKQAMTRSLAALKNM 643
DB 769 LWDA-TSANERKSINVQKQ-FLNLEDPQEDMEVIVKCCSWSADG-----ARIMVAAKN- 819

QY 644 AHKLQTLATNLASEASDKGNPKYSHNSLMVQAIKMTLTDPP----- 687
DB 820 ---KIFLWNTDSRSKVADCRGHL-SWVHGVMFSPDGSSFLTSSDDQTIRLWETKVCYKNS 875

QY 688 -----IHVLPFDVREALI-----IQLLTEASRPNTALISPENLQK 722
DB 876 AVMLKQEVDFVQENVWVLAVDHIRRLQINGRTQIDYLTE--AQVSCCCLSPHLQYI 933

QY 723 ASGSDKGGSF-----LTGKRAAVLFQ--QVKET-----IKENIKEHLDDDEEDEE----- 767
DB 934 AFG--DENGAIETELVNNR---IFQSRFQHKTVWHIQTADAETKLISSSDDAEIQVWN 988

QY 768 -----IMQRRESDEPERSKSLPLTLLYNLTMDTAKLFMSCLHAWGL-----NEVLDE 818
DB 989 WQLDKCIPLRGHOETVKDFRLLKNS--RLLSWS-----FDGTVKVWNITGNKEKDF 1038

QY 819 VCLDRGLMKLPHCTVSGLLSRGHSMLPGYNQACKLSHGKTEVGRK----- 868
DB 1039 VC-----HOGTVL-----SCDISHDATFESSTADKTAIWS 1070

QY 869 -----LPASEGVKGTVGSRAVTTQHLLSIISLANTLMGMTNATFIDGHMKKGTPRPPR 924
DB 1071 FDLLLPLHE--LRGHNGVCAC-----SAFVSDSTLLA-----TGD----- 1103

QY 925 STPDLSKARSGPTSSNIVGOIQKQVNAAPVVSARSADHSG 965
DB 1104 ---DNGEIR-----INWVNSGELLHLCAP-LSBEGAATHGG 1135

```
Db 574 SEVYQAKLQKQEDVNGMLYLEWINKNITN-----LSRLVVRPH-----TDA 617
Qy 413 VYIPAGRLVCGREDGS-IVIVPATOT-AIVQLQGEHMLRGPHTLRGHRNKVTCL 470
Db 618 VY---H---ACFSEDGQRIASCADKTLQVFKATGKLE-----IKAHEDEVLC 663
Qy 471 LYPHQVSARVDQRYLSGGVDFSVIIWDIFSGEMKHLFCVHGGEITQLVPPENC---SA 527
Db 664 AF-----STDDRFIATCSVDKKIKWNSMTGLVHTYDEHSQV-----NCCHFTN 709
Qy 528 RVQHCISV-ASDHSVGLLSREKCKICMLASRLFPQIVIKWRPDDYLWVGS-DGSVY 585
Db 710 SSHLLLATSSDCFLKDLNCKECENTWFGHTNSVNHCRFSP-DDKLLASCSAGTLYK 768
Qy 586 VQWMDTGALDRCVMGITAVEILNACD--EAVPAADVSLSHPAVNLKQAMTRRSALAKNM 643
Db 769 LWDA-TSANERKSINVKQF-FLNLEDPQEDMEVIVKCCSWADG-----ARIMVAAKN- 819
Qy 644 AHKILQTLATNLASEASDKGNLPKYSHNSLWQAIKNLTDPD----- 687
Db 820 ---KIFLWNTDSRSKVADCRGHL-SWVHGVMPSPDGSSPLTSSDDQTIRLWETKVKCKNS 875
Qy 688 ---IHLFFDVEALI-----IQLLTEEASRPNTALISPENLQK 722
Db 876 AVMLKQEVVDVVFQENEVWVLAVDHIRRLQINGRTGQIDYLTE--AQVSCCCLSPHLQYI 933
Qy 723 AGSSDKGGSF-----LTGKRAAVLFQ---QVKET---IKENIKEHLDDDEEDEE----- 767
Db 934 AFG--DENGAIETELVNNR---IFQSRFQHKKTVMHIQFTADEKTLISSDDAEIQVWN 988
Qy 768 -----IMQRRESDEPEYSSKSLPTLLEYNLTMDTAKLFMSCLHAWGL---NEVLDE 818
Db 989 WOLDKCIFLRGHOETVKDFELLKNS--RLLSWS-----FDGTVKVMNIITGNKEKDF 1038
Qy 819 VCLDLGLMKPCHCTVSGLLSRGGMSLMLPGYNQPACKLSHGKTEVGRK----- 868
Db 1039 VC-----HOGTVL-----SCDISHDATKPSSTSDAKTKAIWS 1070
Qy 869 ----LPASGVGKGTGVSRVAVTQHLLSIISLANTLMSMTNATFICDHMKKGTPTRPPR 924
Db 1071 FDLALLPLHE--LRGHNGCYRC-----SAFVSVDSTLLA-----TGD----- 1103
Qy 925 STPDLSKARGSPPTSSNIVOGQIKQVAPVWSARSADHSG 965
Db 1104 ---DNGEIR-----IWNVSNGLLEHLHCAP-LSBEGAATHGG 1135
```

RESULT 15

```
US-08-988-856B-4
; Sequence 4, Application US/0898856B
; Patent No. 6291642
; GENERAL INFORMATION:
; APPLICANT: Weinstein, Jasminder
; TITLE OF INVENTION: No. 6291642el Mammalian Cell Cycle Protein
; FILE REFERENCE: 06843.0026-04; A-283 D
; CURRENT APPLICATION NUMBER: US/08/988,856B
; CURRENT FILING DATE: 1997-12-11
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-988-856B-4
```

```
Query Match 1.9%; Score 150.5; DB 3; Length 499;
Best Local Similarity 19.5%; Pred. No. 6.1e-05;
Matches 86; Conservative 68; Mismatches 160; Indels 126; Gaps 15;
Qy 997 DLGLDKFRPPLLEMLARWQDRCLVREAAQALLLAELRLIEQAGKEAIDAWAPYLPQ 1056
Db 9 DLHSLQLDAPINAPPARWQKAKEAGAPSPMPRA-ANRSHSAGRTFG----- 57
```

```
Qy 1057 YIDHVISPGVTSEAAQTITTAP--DASGPEAKVQEBEHLVDVDDITTCGLSSVPQMKKIS 1114
Db 58 -----RTPGKSSSKVQTTPTSPKPGDDRYIPIHRSAAQME-----VASFLLSKENQSENSQ 105
Qy 1115 TSVEERRKQATAIVLLGVIGABFGAIEPPKLLTRPRSSSQIPEGFG-----LTSGGSNY 1169
Db 106 TPTKKEHQXAWALNMGF-----DVEERAKILRLSGKQONAPEGYQNRILKVLYSOKATP 158
Qy 1170 SLARHTCKALTFLLLQPPSPKLPHPHSTIRRTAIDLIGRGFTVWEPTYMDVSAVLMLLELC 1229
Db 159 GSSRKTCRYI-----PSLPDR-----ILDAPEIR 182
Qy 1230 ADAEKQLANITWGLPLSPAADS-----ARSARHALSLIATARPPAFTITIAKEVHRHTAL 1284
Db 183 NDYYLNLVDWSSGNVLAVALDNSVYLWSASGGDILQLQMEQPGYISSVA-WIKEGNYL 241
Qy 1285 AANTOSQONMHTTTTLARAKGEILRVIEILIEKMTDVLVVEVMDIIMYCLSGSLVKKK 1344
Db 242 AVGTSSAE-----VOLWDV-----QQQK 259
Qy 1345 GLQECFPAICRFYVMSYERNHRIAAGARHGSVALYDITGK--COTIHGHKGPITAVAF 1402
Db 260 RLRRNMTSHSARVGSLSW--NSYLSSGSRSGHHHHDVRAEHHVATLSGHSQEVCGLRW 317
Qy 1403 APDGRYLATYSNTDISHISFW 1422
Db 318 APDGRHLASGNG-DNLVNVW 336
```

Search completed: June 30, 2005, 17:25:20
Job time : 53 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 00:23:10 ; Search time 2091 Seconds
(without alignments)
12663.320 Million cell updates/sec

Title: US-10-645-335-1
Perfect score: 4473
Sequence: 1 atggcgggaacagccttgt.....agcaccgcttcagtgtctaa 4473

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4473	100.0	4473	12	ADQ60211 Human rab
2	4469.8	99.9	4548	4	ADL18130 Human TGF
3	4460.2	99.7	4517	6	ABV99403 Human NOV
4	3549.8	79.4	4497	4	ADL18131 Mouse GF-
5	3537.8	79.1	4494	4	ADL18129 Rat TGF-b
6	3497	78.2	6224	8	ABX70864 Novel hum
7	2807.4	62.8	3316	10	ADA53749 Human cod
8	569.2	12.7	834	5	ADM19384 Novel hum
9	557.6	12.5	630	5	ADM19354 Novel hum
10	474.6	10.6	2578	11	ADM02366 Human cDN
11	393	8.8	4578	13	ADS96553 Drosophil
12	393	8.8	5221	4	ABL19261 Drosophil
13	393	8.8	8411	4	ABL19260 Drosophil
14	393	8.8	13668	4	ABL06236 Drosophil
15	191.8	4.3	491	5	ADM19627 Novel hum
16	150.2	3.4	585	12	ACH74667 Human gen
17	147.2	3.3	227	12	ACH88367 Human gen
18	103.8	2.3	501	12	ACH75227 Human gen
19	103	2.3	356	3	AAC10429 Human sec
20	99	2.2	99	12	ACH88927 Human gen

21	88.4	2.0	555	12	ACH74845	Ach74845 Human gen
22	84	1.9	84	12	ACH88545	Ach88545 Human gen
23	69.2	1.5	3493	12	ADQ63583	Adq63583 Novel hum
24	57	1.3	65	6	ABN30858	Abn30858 Rat eplc
25	48.8	1.1	2000	8	ADA71938	Ada71938 Rice gene
26	48	1.1	2152	2	AAQ70732	AAQ70732 TATA-bind
27	48	1.1	2152	2	AAQ70732	AAQ70732 TATA-bind
28	48	1.1	2152	2	AAQ70732	AAQ70732 TATA-bind
29	48	1.1	3496	10	ADF81713	Adf81713 Leukaemia
30	46.4	1.0	2000	8	ADA71938	Ada71938 Rice gene
31	41.8	0.9	852	13	ADT44265	Adt44265 Bacteria
32	40.8	0.9	1159	6	ABQ68518	Abq68518 Listeria
33	40.8	0.9	1395	6	ABQ70546	Abq70546 Listeria
34	40.6	0.9	774	10	ADE28694	Adg28694 Human NOV
35	40.6	0.9	774	12	ADM93439	Adm93439 Human NOV
36	40.6	0.9	1110	12	ADQ85055	Adg85055 Human tum
37	40.6	0.9	1110	13	ADQ87397	Adg87397 Human tum
38	40.6	0.9	1110	13	ACN38886	Acn38886 Tumour-as
39	40.6	0.9	2862	13	ADQ85056	Adg85056 Human tum
40	40.6	0.9	2862	13	ADQ87398	Adg87398 Human tum
41	40.6	0.9	2862	13	ACN38887	Acn38887 Tumour-as
42	40.6	0.9	2863	4	AAH16061	Aah16061 Human cDN
43	40.6	0.9	3023	4	AAH14005	Aah14005 Human cDN
44	40	0.9	1296	5	AAS70486	Aas70486 DNA encod
45	39.6	0.9	514	12	ACH71629	Ach71629 Human gen

ALIGNMENTS

RESULT 1

ADQ60211

ID ADQ60211 standard; DNA; 4473 BP.

AC ADQ60211;

XX 07-OCT-2004 (first entry)

DE Human rabconnectin 3-binding protein coding sequence.

KW human; rabconnectin 3; GDP/GTP exchange reaction promotion protein;

KW binding protein; calcium dependent exocytosis; gene; db.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 1..4473

FT /*tag= a

FT /product= "Human rabconnectin 3 and GDP/GTP exchange

FT reaction promotion protein-binding protein"

XX JP2004201673-A.

XX 22-JUL-2004.

XX 13-AUG-2003; 2003JP-00207500.

XX 01-NOV-2002; 2002JP-00319521.

XX (EISA) EISAI CO LTD.

XX WPI; 2004-556370/54.

XX P-PSDB; ADQ60212.

XX Novel protein that couples with rabconnectin 3 and GDP/GTP exchange

XX reaction promotion protein, useful for elucidating chemical substance

XX that activates or inhibits calcium dependent exocytosis.

XX Claim 4; SEQ ID NO 1; 52pp; Japanese.

XX The invention comprises the amino acid and coding sequence of a protein

XX that couples with rabconnectin 3 and GDP/GTP exchange reaction promotion

XX protein. The DNA and protein sequences of the invention are useful for

CC analysing immune tissue for locating the protein of the invention or for
CC analysing the expression level of the protein of the invention. The DNA
CC and protein sequences of the invention are useful for elucidating
CC chemical substances that activate or inhibit calcium dependent
CC exocytosis. The present DNA sequence encodes the human rabconnectin 3 and
CC GDP/GTP exchange reaction protein-binding protein of the
CC invention.

AA
SQ Sequence 4473 BP; 1226 A; 1024 C; 1047 G; 1176 T; 0 U; 0 Other;
Query Match 100.0%; Score 4473; DB 12; Length 4473;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4473; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY	1	ATGGCAGGAAA	CAGCCTTGTTCT	CTACCCATPTGTTCTTTTGGGGT	CGAAAAGCGGCCACACAT	60
DB	1	ATGGCAGGAAA	CAGCCTTGTTCT	CTACCCATPTGTTCTTTTGGGGT	CGAAAAGCGGCCACACAT	60
QY	61	TGCATCTCAGCGGTACTTTTAA	CAGATGATGGGCCACGATCGTAA	CAGATGTACGAC	120	
DB	61	TGCATCTCAGCGGTACTTTTAA	CAGATGATGGGCCACGATCGTAA	CAGATGTACGAC	120	
QY	121	GGACAAATATGTCTCTGGGATCTTT	CAGTAGAACATGC	CAAAATTAATCCTCGAGCATGTGTG	180	
DB	121	GGACAAATATGTCTCTGGGATCTTT	CAGTAGAACATGC	CAAAATTAATCCTCGAGCATGTGTG	180	
QY	181	TTTGGTGCATACAGCATCAATCACT	TGTTGTCTTAAAGCTTGTCCTC	CAGTGCACAAACAG	240	
DB	181	TTTGGTGCATACAGCATCAATCACT	TGTTGTCTTAAAGCTTGTCCTC	CAGTGCACAAACAG	240	
QY	241	TATATTTGAGTGCATCTGAAAGTGAGAGATGTG	CGCTCTGGGATGTGAGTATGGCAGA	300		
DB	241	TATATTTGAGTGCATCTGAAAGTGAGAGATGTG	CGCTCTGGGATGTGAGTATGGCAGA	300		
QY	301	TGTATTTGAATTTACAAATTTAGCTTTG	CACACATATCTGGCATACAGTTTAC	CCAGTTCCT	360	
DB	301	TGTATTTGAATTTACAAATTTAGCTTTG	CACACATATCTGGCATACAGTTTAC	CCAGTTCCT	360	
QY	361	GTTTGGGAATCAGCGAGAGAAAGGCTTTTATG	CGCAGGACATTTACCTGAAATCCTTGTT	420		
DB	361	GTTTGGGAATCAGCGAGAGAAAGGCTTTTATG	CGCAGGACATTTACCTGAAATCCTTGTT	420		
QY	421	GTGGATGCTACAGGCTTTGAAGTATTA	TACTCCTTAGTATCAAGATATCA	CCAGACTGG	480	
DB	421	GTGGATGCTACAGGCTTTGAAGTATTA	TACTCCTTAGTATCAAGATATCA	CCAGACTGG	480	
QY	481	ATTAGCTCCATCAGTATTATTTCGATCC	CCAGGACACAGAGAGACACAGTGTGACACTC	540		
DB	481	ATTAGCTCCATCAGTATTATTTCGATCC	CCAGGACACACAGAGAGACACAGTGTGACACTC	540		
QY	541	TGGGTGACTGGCATCCTGAAGTCTGGATTGT	TACTCGGAAATAGTGACATGCAGGAT	600		
DB	541	TGGGTGACTGGCATCCTGAAGTCTGGATTGT	TACTCGGAAATAGTGACATGCAGGAT	600		
QY	601	ACTGAGCCNAATTTTGAGGAGGAATCCAA	CCAAATTTATTTGCAGNAATTTGCCAAGCAATC	660		
DB	601	ACTGAGCCNAATTTTGAGGAGGAATCCAA	CCAAATTTATTTGCAGNAATTTGCCAAGCAATC	660		
QY	661	TCCTTTTGTGCATTTACACAAAGGTCACT	TTTTTGTGTGTGTTTCCAAATATTGGAGGGTG	720		
DB	661	TCCTTTTGTGCATTTACACAAAGGTCACT	TTTTTGTGTGTGTTTCCAAATATTGGAGGGTG	720		
QY	721	TTTCGATGCCGAGACATAATTCCTTGTGTG	TTGTTTCAGGTCCTTAGTGAAATGGACAGCATGG	780		
DB	721	TTTCGATGCCGAGACATAATTCCTTGTGTG	TTGTTTCAGGTCCTTAGTGAAATGGACAGCATGG	780		
QY	781	ACCGGGGGGACCTTTTGTTCTCATCAGATA	AAAGTCATCTTTTGGACAGAAATGGGCAAGT	840		
DB	781	ACCGGGGGGACCTTTTGTTCTCATCAGATA	AAAGTCATCTTTTGGACAGAAATGGGCAAGT	840		
QY	841	TATATTTTCAAACTACCTGCCAGTTGCC	TTTCAGCTTAGTGATTTCAATTCGCGAGTGAATG	900		
DB	841	TATATTTTCAAACTACCTGCCAGTTGCC	TTTCAGCTTAGTGATTTCAATTCGCGAGTGAATG	900		

QY 1981 TCTGACAGGGAATTTACCTAAATATTTCTCATAACTCCCTGATGTTTCAAGCAATAAG 2040
DB 1981 TCTGACAGGGAATTTACCTAAATATTTCTCATAACTCCCTGATGTTTCAAGCAATAAG 2040
QY 2041 ACAAACTTAACAGACCCGGACATACATGCTATTTCTTCTGATGGAAGGTTGATTAT 2100
DB 2041 ACAAACTTAACAGACCCGGACATACATGCTATTTCTTCTGATGGAAGGTTGATTAT 2100
QY 2101 CAATCTCTGATGAAGAAGCTCTAGGCCGAATACTGCTCTTATTTCTCCCGAGAGAATTTG 2160
DB 2101 CAATCTCTGATGAAGAAGCTCTAGGCCGAATACTGCTCTTATTTCTCCCGAGAGAATTTG 2160
QY 2161 CAATAAGCATCTGGCAGTTGACAAAGGGGGCTCTTTTAACTGGAACACGACGCA 2220
DB 2161 CAATAAGCATCTGGCAGTTGACAAAGGGGGCTCTTTTAACTGGAACACGACGCA 2220
QY 2221 GTTCTCTTCCAAAGTGAAGAAGCAATCAAGAGAAATCAAGGAACACCTCTCTTAT 2280
DB 2221 GTTCTCTTCCAAAGTGAAGAAGCAATCAAGAGAAATCAAGGAACACCTCTCTTAT 2280
QY 2281 GATGAAGAGGAGATGAGGAGATAATGAGGCAGAGAAAGGAAGATGATCTCTGAATAT 2340
DB 2281 GATGAAGAGGAGATGAGGAGATAATGAGGCAGAGAAAGGAAGATGATCTCTGAATAT 2340
QY 2341 CGGTCCAGCAAAATCAAAAGCCATTGACCCCTATTAGAAATATAATTTAACTATGACACTGCA 2400
DB 2341 CGGTCCAGCAAAATCAAAAGCCATTGACCCCTATTAGAAATATAATTTAACTATGACACTGCA 2400
QY 2401 AAGCTGTTTATGCTCCCTTCAAGCTGAGCTGGGTTTGAATCAAGTACTGGAATGAATTTG 2460
DB 2401 AAGCTGTTTATGCTCCCTTCAAGCTGAGCTGGGTTTGAATCAAGTACTGGAATGAATTTG 2460
QY 2461 CTGGATCGCTTGGAAATGCTGAACCCCACTGACACCGTATCGTTTGGCCCTCTTGTCAAGA 2520
DB 2461 CTGGATCGCTTGGAAATGCTGAACCCCACTGACACCGTATCGTTTGGCCCTCTTGTCAAGA 2520
QY 2521 GGAGGCCATATGTCATGATGCTGCCGGTTTAAATCAGCGTCTGTTGTAATACTGTACAT 2580
DB 2521 GGAGGCCATATGTCATGATGCTGCCGGTTTAAATCAGCGTCTGTTGTAATACTGTACAT 2580
QY 2581 GGGAAACAGAGTGAAGAGAGCTGCCAGCTCTGAGGAGTGAAGAGGAGACTTAC 2640
DB 2581 GGGAAACAGAGTGAAGAGAGCTGCCAGCTCTGAGGAGTGAAGAGGAGACTTAC 2640
QY 2641 GGAAGTCCCGTCCGTCACACACAGCATCTCTGCTATCATTTCTTTGGCAAAATACT 2700
DB 2641 GGAAGTCCCGTCCGTCACACACAGCATCTCTGCTATCATTTCTTTGGCAAAATACT 2700
QY 2701 TTAATGAGTATGACCAATGCACTTTTATTTGGTGATCATATGAAGAAGGTCCTACAGG 2760
DB 2701 TTAATGAGTATGACCAATGCACTTTTATTTGGTGATCATATGAAGAAGGTCCTACAGG 2760
QY 2761 CCACCTAGACCAAGCACCCAGACCTTTCTAAGCAGGGGTTCCCTCCAACTTCCAGT 2820
DB 2761 CCACCTAGACCAAGCACCCAGACCTTTCTAAGCAGGGGTTCCCTCCAACTTCCAGT 2820
QY 2821 AATATTGTCAAGGACAGATTAAACAGTTGCTGACCTGCTGTTTCCGCTCGCTCTGAT 2880
DB 2821 AATATTGTCAAGGACAGATTAAACAGTTGCTGACCTGCTGTTTCCGCTCGCTCTGAT 2880
QY 2881 GCTGATCACTCTGGCTCTGACCCCTCTTCTGCTCTGCTTTTACATCTGTTCTTAGTA 2940
DB 2881 GCTGATCACTCTGGCTCTGACCCCTCTTCTGCTCTGCTTTTACATCTGTTCTTAGTA 2940
QY 2941 AATGAAGTTGAGTCAAGTCTGATGCTATGCACTGTTGTTATGCTGCCAGACCTACTGGGA 3000
DB 2941 AATGAAGTTGAGTCAAGTCTGATGCTATGCACTGTTGTTATGCTGCCAGACCTACTGGGA 3000
QY 3001 TTGGATAAATTTAGGCTCCCTTCTCGAGATGCTGCCCGAGAGATGGCAAGATCGATGC 3060
DB 3001 TTGGATAAATTTAGGCTCCCTTCTCGAGATGCTGCCCGAGAGATGGCAAGATCGATGC 3060
QY 3061 TTGGAGGTGAGAGAAGCGCACAGGCCCTCTGCTTCTGCGGAACTGAGAGAATTGAGCAG 3120

DB 3061 TTGGAGGTGAGAGAAGCGCACAGGCCCTCTGCTTCTGCGGAACTGAGAGAATTGAGCAG 3120
QY 3121 GCAGGAGAGAAGAACCCATTTGATGCTGGGCTCTTACTTACTCTCAGTACATAGACCAC 3180
DB 3121 GCAGGAGAGAAGAACCCATTTGATGCTGGGCTCTTACTTACTCTCAGTACATAGACCAC 3180
QY 3181 GTCATATCATCTGGAGTACATCAGAAAGCGCGCAGCATATACACAGGCTCTCTGATGCC 3240
DB 3181 GTCATATCATCTGGAGTACATCAGAAAGCGCGCAGCATATACACAGGCTCTCTGATGCC 3240
QY 3241 TCAGGCGCTGAAGCAAAAGTCCAGGAGGAAGCATGACCTTCTGACGATGACATCACC 3300
DB 3241 TCAGGCGCTGAAGCAAAAGTCCAGGAGGAAGCATGACCTTCTGACGATGACATCACC 3300
QY 3301 ACTGGTTGCTTATCAAGTGTCCCAAAATGAAAAAATTTCTACATCTTACGAGGAAGA 3360
DB 3301 ACTGGTTGCTTATCAAGTGTCCCAAAATGAAAAAATTTCTACATCTTACGAGGAAGA 3360
QY 3361 CGGAACAAAGCTACCCGCTATTTGTTTACTTTGAGTAAATAGGAGCTGAAATTTGGTGTGAA 3420
DB 3361 CGGAACAAAGCTACCCGCTATTTGTTTACTTTGAGTAAATAGGAGCTGAAATTTGGTGTGAA 3420
QY 3421 ATTGAACCTCCTAACTATTTGACGACACCTCGAGCTCTAGCCAAATTTCTGAGGGAATTC 3480
DB 3421 ATTGAACCTCCTAACTATTTGACGACACCTCGAGCTCTAGCCAAATTTCTGAGGGAATTC 3480
QY 3481 GGGTTGACTAGTGGTGGATCCCAACTACTCGCTGGCCAGACATCTTGCAGGCACTGACG 3540
DB 3481 GGGTTGACTAGTGGTGGATCCCAACTACTCGCTGGCCAGACATCTTGCAGGCACTGACG 3540
QY 3541 TTTCTTCTGCTACAGCCTCCAAAGCCCAAACTTCTCCACACAGCACTATCCGAAGAACA 3600
DB 3541 TTTCTTCTGCTACAGCCTCCAAAGCCCAAACTTCTCCACACAGCACTATCCGAAGAACA 3600
QY 3601 GCCATTGATCTGATTTGAGCGTGGGTTTCACTGTTTGGAGGCTTACATGATGTTCCGCT 3660
DB 3601 GCCATTGATCTGATTTGAGCGTGGGTTTCACTGTTTGGAGGCTTACATGATGTTCCGCT 3660
QY 3661 GTTCTCATGGGGCTTCTCGAACTTGTGCGGATGCCGAGAAACAACTTGCACACATCACA 3720
DB 3661 GTTCTCATGGGGCTTCTCGAACTTGTGCGGATGCCGAGAAACAACTTGCACACATCACA 3720
QY 3721 ATGGGTTGCTCTGAGCCAGCAGCTGACTCGGCCGCTCTCGAGGCAATGCCCTCTCG 3780
DB 3721 ATGGGTTGCTCTGAGCCAGCAGCTGACTCGGCCGCTCTCGAGGCAATGCCCTCTCG 3780
QY 3781 CTCATTGCCACCGCAGACCAACCGCTTTCATCACCACCATAGCCAAAGGTTACACAGA 3840
DB 3781 CTCATTGCCACCGCAGACCAACCGCTTTCATCACCACCATAGCCAAAGGTTACACAGA 3840
QY 3841 CATACGGCTCTTGCAGCAAAATACCAATCACAGCAGAAATATGCAACAACTCTTGCA 3900
DB 3841 CATACGGCTCTTGCAGCAAAATACCAATCACAGCAGAAATATGCAACAACTCTTGCA 3900
QY 3901 CGAGCTAAAGGGGAAATTTTGGAGTCAATTGAAATTTCTTATTTGAAAAGATGCCACAGAT 3960
DB 3901 CGAGCTAAAGGGGAAATTTTGGAGTCAATTGAAATTTCTTATTTGAAAAGATGCCACAGAT 3960
QY 3961 GTTGTGGATCTTCTGTTGAGGTTATGGACATCATTTATGTTACTGCTTGAAGGATCTTTA 4020
DB 3961 GTTGTGGATCTTCTGTTGAGGTTATGGACATCATTTATGTTACTGCTTGAAGGATCTTTA 4020
QY 4021 GTTAAAAGAAAGGCTCTCAAGAAATGTTTCCAGCCATCTGCAAGTTCTACATGTTGTCAGC 4080
DB 4021 GTTAAAAGAAAGGCTCTCAAGAAATGTTTCCAGCCATCTGCAAGTTCTACATGTTGTCAGC 4080
QY 4081 TATTATGAGCGGAATCACAGAAATAGCAGTTGGAGCTCGCCATGTTTCACTGAGGCTCTGATC 4140
DB 4081 TATTATGAGCGGAATCACAGAAATAGCAGTTGGAGCTCGCCATGTTTCACTGAGGCTCTGATC 4140
QY 4141 GACATCCGAGCTGGAATAATGTTCAGAACTTCATGGAACAGGGAACAACTCATCTGAGTG 4200

Db 4141 GACATCGGACTGGAAAATGTCAGACAATCCATGAGACAAAGGACCAATCACTGCGAGTG 4200
Qy 4201 GCTTTTGTCTCTGATGGAAGATATCTTGCCACTACTCAAACTACACAGCCACATTTCT 4260
Db 4201 GCTTTTGTCTCTGATGGAAGATATCTTGCCACTACTCAAACTACACAGCCACATTTCT 4260
Qy 4261 TTTTGGCAGATGAACAGCTCACTGCTGGGAAGCATCGGATGCTGAACTCGGCACCTCAG 4320
Db 4261 TTTTGGCAGATGAACAGCTCACTGCTGGGAAGCATCGGATGCTGAACTCGGCACCTCAG 4320
Qy 4321 CTGCGTGCATTAAACCTACCAAGTGCCCTCTGTGAGCCCGCTGCCCGCTCCAC 4380
Db 4321 CTGCGTGCATTAAACCTACCAAGTGCCCTCTGTGAGCCCGCTGCCCGCTCCAC 4380
Qy 4381 AATGCCCTCAAGCTGGCCGGCTCATCTGCACTTCCAAACCGCAACCTCATCTCATGGCC 4440
Db 4381 AATGCCCTCAAGCTGGCCGGCTCATCTGCACTTCCAAACCGCAACCTCATCTCATGGCC 4440
Qy 4441 CATGACGGGAAGGAGCACCGCTTCATGCTCTAA 4473
Db 4441 CATGACGGGAAGGAGCACCGCTTCATGCTCTAA 4473

RESULT 2

AAD18130

ID AAD18130 standard; cDNA; 4548 BP.

XX AC

XX AAD18130;

XX 18-DEC-2001 (first entry)

XX Human TGF-beta Resistance Associated Gene (TRAG) cDNA.

XX Human; Transforming growth factor-beta; TGF-beta; cellular proliferation;

XX phenotype; TRAG; TGF-beta Resistance Associated Gene; cancer therapy;

XX oncogenesis; gene mapping; transgenic animal; chromosome 18; ss.

XX Homo sapiens.

XX Key

XX Location/Qualifiers

XX 76..4548

XX /*tag= a

XX /product= "Human TRAG protein"

XX W0200166739-A1.

XX 13-SEP-2001.

XX 12-FEB-2001; 2001WO-US004475.

XX 07-MAR-2000; 2000US-0187572P.

XX 08-MAR-2000; 2000US-0187848P.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Thorgeirsson SS, Sanders S;

XX WPI; 2001-589942/66.

XX P-PSDB; AAE10802.

XX Novel isolated polypeptide for studying and modulating mechanisms

XX involved in cellular proliferation comprises transforming growth factor-

XX beta resistance associated polypeptide fragment.

XX Claim 7; Page 56-61; 89pp; English.

XX The present sequence is a cDNA encoding human transforming growth factor

XX (TGF)-beta resistance associated (TRAG) protein. Human TRAG gene is

XX located on chromosome 18. TRAG DNA or protein is useful for studying and

XX modulating mechanisms involved in cellular proliferation, and for

XX modulating cellular phenotype. It is also used for screening aggressive

XX metastasis in cancer cells, and for targets used in cancer therapy. TRAG

XX protein is useful for evaluating factors that interact with and/or

CC control TGF-beta signalling for understanding both cell proliferation
CC control and oncogenesis, and in protein-protein interaction cells. TRAG
CC DNA is useful as hybridisation probe, in chromosome and gene mapping, in
CC the generation of antisense RNA and DNA, for the preparation of TRAG
CC protein by recombinant techniques, and for generating either transgenic
CC or knockout animals which, in turn, are useful in the development and
CC screening of therapeutically useful reagents

XX Sequence 4548 BP; 1255 A; 1035 C; 1056 G; 1202 T; 0 U; 0 Other;

SQ Query Match 99.9%; Score 4469.8; DB 4; Length 4548;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 4471; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCAGGAAACAGCGCTTCTTCTACCCATTGTTCTTTGGGTCGAAAAGCGCCACACAT 60

Db 76 ATGGCAGGAAACAGCGCTTCTTCTACCCATTGTTCTTTGGGTCGAAAAGCGCCACACAT 135

Qy 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAAACAGGATGTCCAGAC 120

Db 136 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAAACAGGATGTCCAGAC 195

Qy 121 GGCAAAATATGCTCTGGGATCTTTTAACTAGATGATGGGGCCACGATCGTAAACAGGATGTCCAGAC 180

Db 196 GGCAAAATATGCTCTGGGATCTTTTAACTAGATGATGGGGCCACGATCGTAAACAGGATGTCCAGAC 255

Qy 181 TTTGGTCATACAGCATCAATCACTTGTGTTTGTCTTAAAGCTTGTCTTCCAGTCGAAAACAG 240

Db 256 TTTGGTCATACAGCATCAATCACTTGTGTTTGTCTTAAAGCTTGTCTTCCAGTCGAAAACAG 315

Qy 241 TATATTGTGAGTGCATCTGAAAAGTGGAGAGATGTGCTCTGGGATGTGAGTGTGAGTGTGAGTGTG 300

Db 316 TATATTGTGAGTGCATCTGAAAAGTGGAGAGATGTGCTCTGGGATGTGAGTGTGAGTGTGAGTGTG 375

Qy 301 TGATTGAATTTACAAAATTTAGCTTGGCAACATATCTGGCATACAGTTCTTACAGTTCTCT 360

Db 376 TGATTGAATTTACAAAATTTAGCTTGGCAACATATCTGGCATACAGTTCTTACAGTTCTCT 435

Qy 361 GTTGGGAATCAGGAGAGAGAGGCTTTTATGCGAGGACATTTACCGCTCGAAATCTCTTGT 420

Db 436 GTTGGGAATCAGGAGAGAGAGGCTTTTATGCGAGGACATTTACCGCTCGAAATCTCTTGT 495

Qy 421 GTGGATGCTACCGAGCTTGAAGTATTATATCTCTTAGTATCAAAAGATATCAACAGACTGG 480

Db 496 GTGGATGCTACCGAGCTTGAAGTATTATATCTCTTAGTATCAAAAGATATCAACAGACTGG 555

Qy 481 ATTAGCTCATGAGTATTATTCGATCCACCGAACAAGAGGACACAGTGTGAGTGTGAGTGTG 540

Db 556 ATTAGCTCATGAGTATTATTCGATCCACCGAACAAGAGGACACAGTGTGAGTGTGAGTGTG 615

Qy 541 TCGGTGACTGGCATCTGAGGCTTGGATTGTTTACCTCGGAAATAGTGCATGCAAGGAT 600

Db 616 TCGGTGACTGGCATCTGAGGCTTGGATTGTTTACCTCGGAAATAGTGCATGCAAGGAT 675

Qy 601 ACTGAGCCAAATATTTGAGGAGGAATCCAAACCAATTTATTGTTCAGAAATGCCAAAGCATC 660

Db 676 ACTGAGCCAAATATTTGAGGAGGAATCCAAACCAATTTATTGTTCAGAAATGCCAAAGCATC 735

Qy 661 TCTTTTGTGCAATTTACAAAGGTCATCTTTGGTGTGTGTTTCCAAATATTTGGAGGGTG 720

Db 736 TCTTTTGTGCAATTTACAAAGGTCATCTTTGGTGTGTGTTTCCAAATATTTGGAGGGTG 795

Qy 721 TTCGATGCCGGAGACTATTCTTGTGTTGTTTACGCTTAGTGAATAATGACAGACATGG 780

Db 796 TTCGATGCCGGAGACTATTCTTGTGTTGTTTACGCTTAGTGAATAATGACAGACATGG 855

Qy 781 ACCGGGGGAGACTTTTGTCTCATCAGATAAAGTTCATTTTGGACAGAAAATGGCAAAGT 840

Db 856 ACCGGGGGAGACTTTTGTCTCATCAGATAAAGTTCATTTTGGACAGAAAATGGCAAAGT 915

Qy 841 TATATTACAAACTACCTGCGAGTTCGCTTCCAGCTAGTGAATTCATTCGCGAGTGTG 900

Db 916 TATATTACAAACTACCTGCGAGTTCGCTTCCAGCTAGTGAATTCATTCGCGAGTGTG 975

3136	Db	TTGGAGGTGAGAGAAGCGCACAGGCCCTGCTTCTGGCGGAACTGAGAAGAAATTGAGCAG	3195
3121	Qy	GCAGGCAGGAAGGAGCCATTGATGCTGGGCTCTTACTTACTCTCAGTACATAGACCAC	3180
3196	Db	GCAGGCAGGAAGGAGCCATTGATGCTGGGCTCTTACTTACTCTCAGTACATAGACCAC	3255
3181	Qy	GTCAATATCACTCGAGTACATCAGAAGCGCGCAGACTATCACACGGCTCTCGATGCC	3240
3256	Db	GTCAATATCACTCGAGTACATCAGAAGCGCGCAGACTATCACACGGCTCTCGATGCC	3315
3241	Qy	TCAGGGCCTGAAGCAAAAGTCCAGGAGGAAGAGCATGACCTTGTGTGACGATGACATCACC	3300
3316	Db	TCAGGGCCTGAAGCAAAAGTCCAGGAGGAAGAGCATGACCTTGTGTGACGATGACATCACC	3375
3301	Qy	ACTGGTTGCTTATCAAGTGTCCTCAAAATGAAAAAATTTCTACATCTTACGAGGAAAGA	3360
3376	Db	ACTGGTTGCTTATCAAGTGTCCTCAAAATGAAAAAATTTCTACATCTTACGAGGAAAGA	3435
3361	Qy	CGGAAGCAAGCTACCGCTATTGTTTACTTTGGAGTATAGAGCTGAATTTGCTGCTGAA	3420
3436	Db	CGGAAGCAAGCTACCGCTATTGTTTACTTTGGAGTATAGAGCTGAATTTGCTGCTGAA	3495
3421	Qy	ATTGAAACCTCCTAAACTATTGACCAGACCTCGAAGCTCTAGCCAAATTCCTGAGGGATTCT	3480
3496	Db	ATTGAAACCTCCTAAACTATTGACCAGACCTCGAAGCTCTAGCCAAATTCCTGAGGGATTCT	3555
3481	Qy	GGGTTGACTAGTGTGGANTCAACTACTCGTGGCGCAGACATCTTGTGCAAGGCATCTGACG	3540
3556	Db	GGGTTGACTAGTGTGGANTCAACTACTCGTGGCGCAGACATCTTGTGCAAGGCATCTGACG	3615
3541	Qy	TTTTCTTCTGTCTACAGCTCCAAAGCCCAAACTTCTCCACACAGACATATCCGAAGAACA	3600
3616	Db	TTTTCTTCTGTCTACAGCTCCAAAGCCCAAACTTCTCCACACAGACATATCCGAAGAACA	3675
3601	Qy	GCCATTGATCTGATTGGAACGTGGGTTCACTGTTTGGAGGCTTACATGAGTGTGTCGCT	3660
3676	Db	GCCATTGATCTGATTGGAACGTGGGTTCACTGTTTGGAGGCTTACATGAGTGTGTCGCT	3735
3661	Qy	GTTCTGATGGGCTCTTCGAACTTTGTGGATGCCAGAGAAACAACTTGGCCACATCACA	3720
3736	Db	GTTCTGATGGGCTCTTCGAACTTTGTGGATGCCAGAGAAACAACTTGGCCACATCACA	3795
3721	Qy	ATGGGTTGCTCTGAGCCAGCAGCTGACTCGSCCGCTCTGCGAGGCATGCCCTCTCG	3780
3796	Db	ATGGGTTGCTCTGAGCCAGCAGCTGACTCGSCCGCTCTGCGAGGCATGCCCTCTCG	3855
3781	Qy	CTCATTTGCCACCGCCAGACACCCGCTTTCATACACCACTAGCCAAAGAGGTACACAGA	3840
3856	Db	CTCATTTGCCACCGCCAGACACCCGCTTTCATACACCACTAGCCAAAGAGGTACACAGA	3915
3841	Qy	CATACGGCTCTTGCAGCAAAATACCCAAATTCACAGCAGAAATGCAACAACATCTTTGCA	3900
3916	Db	CATACGGCTCTTGCAGCAAAATACCCAAATTCACAGCAGAAATGCAACAACATCTTTGCA	3975
3901	Qy	CGAGCTAAAGGGGAAATTTTGAGAGTCAATTGAAATTCCTTATTTGAAAGATGCCACAGAT	3960
3976	Db	CGAGCTAAAGGGGAAATTTTGAGAGTCAATTGAAATTCCTTATTTGAAAGATGCCACAGAT	4035
3961	Qy	GTTGTGGATCTTCTCGTGGAGGTTATGGACATCATATTATGTACTGCCCTTGAGGATCTTTTA	4020
4036	Db	GTTGTGGATCTTCTCGTGGAGGTTATGGACATCATATTATGTACTGCCCTTGAGGATCTTTTA	4095
4021	Qy	GTTAAAAAGAAAGTCTTCAAGAAATGTTTCCAGCCATCTGCGAGGTTCTACATGGTCAGC	4080
4096	Db	GTTAAAAAGAAAGTCTTCAAGAAATGTTTCCAGCCATCTGCGAGGTTCTACATGGTCAGC	4155
4081	Qy	TATTATGAGCGGAATCACAGAAATAGCAGTTGGAGCTGCCATGCTTCACTGGGCTCTGAC	4140
4156	Db	TATTATGAGCGGAATCACAGAAATAGCAGTTGGAGCTGCCATGCTTCACTGGGCTCTGAC	4215
4141	Qy	GACATCCGGAATCGGAAAAATGTGACAAATCCATGGACACAGGGAACCAATCACTGCAGTG	4200

Db	4216	GACATCCGAGCTGGAAATGTCTAGACAATCCATGGACACAAAGGACCAATCACTGCACTG	42175
Qy	4201	GCCTTTTGCTCTCTGATGGAAGATATCTTTGGCCACCTACTCTCAAAACACTGACAGCCACATTTCT	4260
Db	4276	GCCTTTTGCTCTCTGATGGAAGATATCTTTGGCCACCTACTCTCAAAACACTGACAGCCACATTTCT	4335
Qy	4261	TTTTGGCAGATGAACACGCTCACTGCTGGGAAGCATCGGCATGCTGAACCTGGGACCTCTCAG	4320
Db	4336	TTTTGGCAGATGAACACGCTCACTGCTGGGAAGCATCGGCATGCTGAACCTGGGACCTCTCAG	4395
Qy	4321	CTGGGCTGCAATTAATAACCTACCAAGGTGCCCGCTGTGTCAGCCCGCTCCCGGCTCCAC	4380
Db	4396	CTGGGCTGCAATTAATAACCTACCAAGGTGCCCGCTGTGTCAGCCCGCTCCCGGCTCCAC	4455
Qy	4381	AATGCCCTCAAGCTGGCCCGGCTCATCTGACACTTCCAAACCGCAACGTCATCTCATGGCC	4440
Db	4456	AATGCCCTCAAGCTGGCCCGGCTCATCTGACACTTCCAAACCGCAACGTCATCTCATGGCC	4515
Qy	4441	CATGACGGGAAGGAGCACCGCTTCATGGTCTAA	4473
Db	4516	CATGACGGGAAGGAGCACCGCTTCATGGTCTAA	4548
RESULT 3			
ABV99403			
ID	ABV99403	standard; DNA; 4517 BP.	
XX	AC	ABV99403;	
XX	XX		
XX	DT	27-JAN-2003 (first entry)	
XX	XX		
DE	DE	Human NOV36a coding sequence.	
XX	XX	Human; anti-HIV; cytostatic; antidiabetic; antiaesthetic; cachexia; AIDS;	
KW	KW	antiinflammatory; cardiant; haemostatic; neuroprotective; anorectic;	
KW	KW	nootropic; immunosuppressive; osteopathic; antiparkinsonian; cancer;	
KW	KW	antifertility; cerebroprotective; gene therapy; NOVX; NOV; fertility;	
KW	KW	metabolic disorder; diabetes; obesity; infectious disease; anorexia;	
KW	KW	neurodegenerative disease; Alzheimer's disease; Parkinson's disease;	
KW	KW	immune disorder; haematopoietic disorder; cardiovascular disorder;	
KW	KW	bronchial asthma; dyslipidemia; metabolic disturbance; neurogenesis;	
KW	KW	metabolic syndrome X; wasting disorder; cell differentiation;	
KW	KW	Single nucleotide polymorphism; SNP; cell proliferation; haematopoiesis;	
KW	KW	wound healing; angiogenesis; gene; ds.	
OS	XX	Homo sapiens.	
XX	XX		
Key	Location/Qualifiers		
FH	variation	replace (91,G)	
FT		/*tag= a	
FT		/standard_name= "Single nucleotide polymorphism"	
FT	variation	replace (230,C)	
FT		/*tag= b	
FT		/standard_name= "Single nucleotide polymorphism"	
FT	variation	replace (4250,G)	
FT		/*tag= c	
FT		/standard_name= "Single nucleotide polymorphism"	
XX	XX		
PN	WO200272771-A2.		
XX	XX		
PD	19-SEP-2002.		
XX	XX		
PF	08-MAR-2002; 2002WO-US007288.		
XX	XX		
PR	08-MAR-2001; 2001US-0274101P.		
PR	08-MAR-2001; 2001US-0274194P.		
PR	08-MAR-2001; 2001US-0274281P.		
PR	08-MAR-2001; 2001US-0274322P.		
PR	09-MAR-2001; 2001US-0274849P.		
PR	12-MAR-2001; 2001US-0275235P.		
PR	13-MAR-2001; 2001US-0275578P.		
PR	13-MAR-2001; 2001US-0275579P.		
PR	13-MAR-2001; 2001US-0275601P.		

PR 14-MAR-2001; 2001US-0276000P.
PR 16-MAR-2001; 2001US-0276776P.
PR 19-MAR-2001; 2001US-0276994P.
PR 20-MAR-2001; 2001US-0277239P.
PR 20-MAR-2001; 2001US-0277321P.
PR 20-MAR-2001; 2001US-0277327P.
PR 20-MAR-2001; 2001US-0277338P.
PR 21-MAR-2001; 2001US-0277791P.
PR 22-MAR-2001; 2001US-0277813P.
PR 23-MAR-2001; 2001US-0278152P.
PR 26-MAR-2001; 2001US-0278894P.
PR 27-MAR-2001; 2001US-0278999P.
PR 28-MAR-2001; 2001US-0279036P.
PR 28-MAR-2001; 2001US-0279344P.
PR 30-MAR-2001; 2001US-0279995P.
PR 30-MAR-2001; 2001US-0280233P.
PR 02-APR-2001; 2001US-0280802P.
PR 02-APR-2001; 2001US-0280822P.
PR 02-APR-2001; 2001US-0280900P.
PR 04-APR-2001; 2001US-0281194P.
PR 13-APR-2001; 2001US-0283675P.
PR 30-APR-2001; 2001US-0287424P.
PR 02-MAY-2001; 2001US-0288066P.
PR 03-MAY-2001; 2001US-0288342P.
PR 03-MAY-2001; 2001US-0288528P.
PR 15-MAY-2001; 2001US-0291190P.
PR 16-MAY-2001; 2001US-0291099P.
PR 16-MAY-2001; 2001US-0291240P.
PR 30-MAY-2001; 2001US-0294485P.
PR 31-MAY-2001; 2001US-0294889P.
PR 31-MAY-2001; 2001US-0294899P.
PR 18-JUN-2001; 2001US-0299027P.
PR 19-JUN-2001; 2001US-0299303P.
PR 19-JUN-2001; 2001US-0299310P.
PR 10-JUL-2001; 2001US-0304354P.
PR 31-JUL-2001; 2001US-0309198P.
PR 16-AUG-2001; 2001US-0312903P.
PR 10-SEP-2001; 2001US-0318462P.
PR 12-SEP-2001; 2001US-0318770P.
PR 27-SEP-2001; 2001US-0325430P.
PR 27-SEP-2001; 2001US-0325681P.
PR 18-OCT-2001; 2001US-0330380P.
PR 31-OCT-2001; 2001US-0335301P.
PR 14-NOV-2001; 2001US-0332172P.
PR 14-NOV-2001; 2001US-0332271P.
PR 14-NOV-2001; 2001US-0332272P.
PR 14-NOV-2001; 2001US-0333184P.
PR 14-NOV-2001; 2001US-0333272P.
PR 21-NOV-2001; 2001US-0332094P.
PR 03-DEC-2001; 2001US-0337426P.
PR 03-DEC-2001; 2001US-0338092P.
PR 04-DEC-2001; 2001US-0337185P.
PR 03-JAN-2002; 2002US-0345705P.
PR 08-MAR-2002; 2002US-00093463.

(CURA-) CURAGEN CORP.

XX Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ;
XX Boldog FI, Li L, Zerhusen BD, Tchernev VT, Gangolli EA, Vernet CAM;
PI Pena CE, Burgess CE, Liu X, Spytek KA, Gorman L, Spaderna SK;
PI Voss EZ, Malyankar UM, Anderson DW, Patturajan M, Miller CE;
PI Taupier RJ, Padigaru M, Shenoy SG, Kekuda R, Gusev VY, Pochart PF;
PI Zhong M;

XX WPI; 2002-732824/79.
DR P-PSDB; ABP70125.

XX New NOVX polypeptides and polynucleotides, useful for preventing,
PT diagnosing or treating NOVX-associated disorders e.g. diabetes, cancer,
PT Alzheimer's disease, dyslipidemias, obesity, immune or hematopoietic
PT disorders, and asthma.

XX Claim 16; Page 230-232; 619pp; English.

XX The present invention relates to new isolated proteins (NOVX) and their
CC coding sequences (ABV99327-ABV99595 and ABP70049-ABP70149), where X is
CC any number from 1 to 48. The NOVX proteins and coding sequences are
CC useful in the manufacture of a medicament for treating a syndrome
CC associated with a human disease. Preferably a NOVX-associated disorder.
CC The NOVX coding sequences and proteins are useful for treating, diabetes,
CC preventing or diagnosing diseases such as metabolic disorders, diabetes,
CC obesity, infectious disease, anorexia, cancer-associated cachexia,
CC cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's
CC disease, immune disorders, hematopoietic disorders, cardiovascular
CC disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic
CC disturbances associated with obesity, metabolic syndrome X or wasting
CC disorders associated with chronic diseases or various cancers. The NOVX
CC coding sequences and proteins may also be used as targets for the
CC identification of small molecules that modulate or inhibit e.g.
CC neurogenesis, cell differentiation, cell proliferation, hematopoiesis,
CC wound healing and angiogenesis, in gene therapy, in generation of
CC antibodies that bind immunospecifically to NOVX substances for use in
CC therapeutic or diagnostic methods
XX

SQ Sequence 4517 BP; 1236 A; 1038 C; 1059 G; 1184 T; 0 U; 0 Other;

Query Match 99.7%; Score 4460.2; DB 6; Length 4517;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 4465; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 ATGGCAGGAACACAGCCTTGTCTTACCCATTGTTCTTTGGGGTGGAAAGCGCCACACAT 60
DB 20 ATGGCAGGAACACAGCCTTGTCTTACCCATTGTTCTTTGGGGTGGAAAGCGCCACACAT 79
QY 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACCGATCTTAACAGATCTCAGCAG 120
DB 80 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACCGATCTTAACAGATCTCAGCAG 139
QY 121 GGCACAAATATGCTCTGGGATCTTTTCAGTAGAAGTCTGCAATTAATCTTCGAGCAGCTGTTG 180
DB 140 GGCACAAATATGCTCTGGGATCTTTTCAGTAGAAGTCTGCAATTAATCTTCGAGCAGCTGTTG 199
QY 181 TTTGGTCATACAGCATCAATCACTGTTGTTGTCTTAAAGCTTGTCTTCCAGTGACAAACAG 240
DB 200 TTTGGTCATACAGCATCAATCACTGTTGTTGTCTTAAAGCTTGTCTTCCAGTGACAAACAG 259
QY 241 TATATTGTGAGTCATCTGAAAGTGGAGAGATGTCCTCTGGGATGTGAGTGTGAGCAGA 300
DB 260 TATATTGTGAGTCATCTGAAAGTGGGAGATGTCCTCTGGGATGTGAGTGTGAGCAGA 319
QY 301 TGTATTGAATTTACAAAATAGCTTGCACACATCTACCTGGCATACAGTTCTACCAAGTTCTCT 360
DB 320 TGTATTGAATTTACAAAATAGCTTGCACACATCTACCTGGCATACAGTTCTACCAAGTTCTCT 379
QY 361 GTTGGGAATCAGCGAGAAGGAGGCTTTTATGCCCGGACATTCACCTGAAATCCTTGT 420
DB 380 GTTGGGAATCAGCGAGAAGGAGGCTTTTATGCCCGGACATTCACCTGAAATCCTTGT 439
QY 421 GTGGATGTACCAGCCTTGAAGTATTATATCTCTTAGTATCAAGATATACACAGACTGG 480
DB 440 GTGGATGTACCAGCCTTGAAGTATTATATCTCTTAGTATCAAGATATACACAGACTGG 499
QY 481 ATTAGTCTCATGAGTATTATTGATCCACCGACACAGAGGACACAGTGTGAGCAGCTC 540
DB 500 ATTAGTCTCATGAGTATTATTGATCCACCGACACAGAGGACACAGTGTGAGCAGCTC 559
QY 541 TCGGTGACTGGCATCTGGAAGTCTGGATTGTTTACCTCGGAATAAGTCACATGCAAGGAT 600
DB 560 TCGGTGACTGGCATCTGGAAGTCTGGATTGTTTACCTCGGAATAAGTCACATGCAAGGAT 619
QY 601 ACTGAGCCCAATATTTTGAGGAGGAATCCAAACCAATTTATTGTGAGATTTGCCAAAGCATC 660
DB 620 ACTGAGCCCAATATTTTGAGGAGGAATCCAAACCAATTTATTGTGAGATTTGCCAAAGCATC 679
QY 661 TCTTTTGTGCAATTTACAAAGGTCATCTTTGTTGTTGTGTTTCCAAATATTGGAGGGTG 720

Db 680 TCTTTTGTGTCATTTTACACAAAGGTCACATTTTGTGTGTGTTCTCCAAATATTGGAGGGTA 739
Qy 721 TTCCGATCCGGAGACATATTCCTTGTGTGTGTTCAGGTCTTAGTGAATAATGGACAGACATGG 780
Db 740 TTCCGATCCGGAGACATATTCCTTGTGTGTGTTCAGGTCTTAGTGAATAATGGACAGACATGG 799
Qy 781 ACCGGGGGGACATTTCTCATCAGATAAAGTCATCATTTTGGACAGAAAATGGCAAAAGT 840
Db 800 ACCGGGGGGACATTTCTCATCAGATAAAGTCATCATTTTGGACAGAAAATGGCAAAAGT 859
Qy 841 TATATTTACAAACTACCTGCAGTTCCTTCCAGCTAGTAGATTCATTCGCCAGTGTG 900
Db 860 TATATTTACAAACTACCTGCAGTTCCTTCCAGCTAGTAGATTCATTCGCCAGTGTG 919
Qy 901 GGGAAAGGCGAGTTGAAAATTTAATTCCTCTGTACACATATCTCTTGGATCGAAAAGAT 960
Db 920 GGGAAAGGCGAGTTGAAAATTTAATTCCTCTGTACACATATCTCTTGGATCGAAAAGAT 979
Qy 961 AAAGAGTTGCTAATTTGTCTCTCTGTACTCTCGGTTCCTTATGGATCGAGAAATATTTTC 1020
Db 980 AAAGAGTTGCTAATTTGTCTCTCTGTACTCTCGGTTCCTTATGGATCGAGAAATATTTTC 1039
Qy 1021 CATAAACTGTATAATTCAGGGTGATTCCTCTGGAAGGTTGAATATTTGGAACATATCAGAC 1080
Db 1040 CATAAACTGTATAATTCAGGGTGATTCCTCTGGAAGGTTGAATATTTGGAACATATCAGAC 1099
Qy 1081 ACAGCTGATAAACAGGGAAGTGAAGAGGCTGGCAATGACAACTTCTATTAGTTGCAA 1140
Db 1100 ACAGCTGATAAACAGGGAAGTGAAGAGGCTGGCAATGACAACTTCTATTAGTTGCAA 1159
Qy 1141 GAGGCAATTTGATAAATCTGATCTCTGTCTGGAATATATAGATCAGCTGAGTGATTT 1200
Db 1160 GAGGCAATTTGATAAATCTGATCTCTGTCTGGAATATATAGATCAGCTGAGTGATTT 1219
Qy 1201 CCCAATAGTAATGAACCTCTTAAAGTAACCTGCAAGTGTGTACATACAGCACATGACCA 1260
Db 1220 CCCAATAGTAATGAACCTCTTAAAGTAACCTGCAAGTGTGTACATACAGCACATGACCA 1279
Qy 1261 CTTGTTTGTGGTCTGTAAGATGAAGCATAGTTATTTGTACCTGCGCACAGACGGCCATA 1320
Db 1280 CTTGTTTGTGGTCTGTAAGATGAAGCATAGTTATTTGTACCTGCGCACAGACGGCCATA 1339
Qy 1321 GTACAGCTGTGCAAGGGGAACACATGCTCAGAAAGGTTGGCCACTCACAGAACACTC 1380
Db 1340 GTACAGCTGTGCAAGGGGAACACATGCTCAGAAAGGTTGGCCACTCACAGAACACTC 1399
Qy 1381 CGTGGTCATCGGAACAAAGTCAATGTTGTCTATATCCTCATCAGGTCTCAGCTCGGTAT 1440
Db 1400 CGTGGTCATCGGAACAAAGTCAATGTTGTCTATATCCTCATCAGGTCTCAGCTCGGTAT 1459
Qy 1441 GATCAAGAATACCTGATATCTGAGGTGTGGATTTTTCAGTCATAATTTGGGACATATTT 1500
Db 1460 GATCAAGAATACCTGATATCTGAGGTGTGGATTTTTCAGTCATAATTTGGGACATATTT 1519
Qy 1501 TCTGGAGAAATGAAACATATCTTCTGTGTTCACTGGTGGTGAGATTAATCAACTTCTAGTT 1560
Db 1520 TCTGGAGAAATGAAACATATCTTCTGTGTTCACTGGTGGTGAGATTAATCAACTTCTAGTT 1579
Qy 1561 CCACCTGAAACCTGTAGTGAAGAGTACAGCACTGATCTGCTGTAGCCAGTGACCA 1620
Db 1580 CCACCTGAAACCTGTAGTGAAGAGTACAGCACTGATCTGCTGTAGCCAGTGACCA 1639
Qy 1621 TCAGTAGGACTTCTAAGTTTGGAGAGAAATAATGCAATATGTTGGCACTCTCGTCACTT 1680
Db 1640 TCAGTAGGACTTCTAAGTTTGGAGAGAAATAATGCAATATGTTGGCACTCTCGTCACTT 1699
Qy 1681 TTTTCTATTCAAGTAATCAAACTGAGGCTTCTGTAGTATTAATGTTGGTGGGGTGTTC 1740
Db 1700 TTTTCTATTCAAGTAATCAAACTGAGGCTTCTGTAGTATTAATGTTGGTGGGGTGTTC 1759
Qy 1741 GATGGTTCTGTGTACGTCTGGCAAAATGGATACTGGTGCAATGGATCGTTGTGTGATGGGG 1800
Db 1760 GATGGTTCTGTGTACGTCTGGCAAAATGGATACTGGTGCAATGGATCGTTGTGTGATGGGG 1819

Qy 1801 ATAAACAGCAGTTGAGATTTCTAAACGCTTGTGATGAAGCTGTTCTCTGCTGCTGTGATTCA 1860
Db 1820 ATAAACAGCAGTTGAGATTTCTAAACGCTTGTGATGAAGCTGTTCTCTGCTGCTGTGATTCA 1879
Qy 1861 CTTAGTCATCCAGCAGTCAACCTTAAACAAAGCTATGAACAGAGCTAGTCTTGTGCTCTTT 1920
Db 1880 CTTAGTCATCCAGCAGTCAACCTTAAACAAAGCTATGAACAGAGCTAGTCTTGTGCTCTTT 1939
Qy 1921 AAAAATATGGCCCATCATTAAGCTACAAACCTTGCACCTAACTCTCTTGGCTTCTGAGGCA 1980
Db 1940 AAAAATATGGCCCATCATTAAGCTACAAACCTTGCACCTTAACTCTTGTGCTTCTGAGGCA 1999
Qy 1981 TCTGCAAGGGGAAATTTTACATAATATCTCATAACTCCCTGATGGTTCACGCAATAAAG 2040
Db 2000 TCTGCAAGGGGAAATTTTACATAATATCTCATAACTCCCTGATGGTTCACGCAATAAAG 2059
Qy 2041 ACAAACCTAAACAGACCCGACATACATGTGCTATTCTTTGATGTGGAGCGTTGATTT 2100
Db 2060 ACAAACCTAAACAGACCCGACATACATGTGCTATTCTTTGATGTGGAGCGTTGATTT 2119
Qy 2101 CAACTCTGACTGAAGAGCCTTAGGCCGAAATACCTGCTTTATTTCCCGCAGAGAAATTTG 2160
Db 2120 CAACTCTGACTGAAGAGCCTTAGGCCGAAATACCTGCTTTATTTCCCGCAGAGAAATTTG 2179
Qy 2161 CAAAAGCATCTGGCAGTTTCAGACAAAGGGGCTCTTTTTTAACTGGAACAAACGAGCAGCA 2220
Db 2180 CAAAAGCATCTGGCAGTTTCAGACAAAGGGGCTCTTTTTTAACTGGAACAAACGAGCAGCA 2239
Qy 2221 GTTCTCTTCCAAACAGTGAAGAAACGATCAAAGAGAAACATCAAGGAAACCTCTCTTGAT 2280
Db 2240 GTTCTCTTCCAAACAGTGAAGAAACGATCAAAGAGAAACATCAAGGAAACCTCTCTTGAT 2299
Qy 2281 GATGAAGAGGAGATGAGGAGATTAATGAGGCGAGAGAGGAGAGAGATGATCCTGATAT 2340
Db 2300 GATGAAGAGGAGATGAGGAGATTAATGAGGCGAGAGAGGAGAGAGATGATCCTGATAT 2359
Qy 2341 CGGTCCAGCAAAATCAAAGCATTTGACCTATTAGATAATAATTTAACTATGACACTGCA 2400
Db 2360 CGGTCCAGCAAAATCAAAGCATTTGACCTATTAGATAATAATTTAACTATGACACTGCA 2419
Qy 2401 AAGCTGTTTATGTCCTGCTTCAAGCTGGGGTTGAAATGAAGTACTGAGTAAGTTTGC 2460
Db 2420 AAGCTGTTTATGTCCTGCTTCAAGCTGGGGTTGAAATGAAGTACTGAGTAAGTTTGC 2479
Qy 2461 CTGGATCGCTTGGATGCTGAAACCCCACTGACCGTATCGTTTGGCTCTTGTCAAGA 2520
Db 2480 CTGGATCGCTTGGATGCTGAAACCCCACTGACCGTATCGTTTGGCTCTTGTCAAGA 2539
Qy 2521 GGAGGCGCATATGTCACCTGATGCTGCGGGTTAATCAGCCTGCTTGTAAACTGTCAACAT 2580
Db 2540 GGAGGCGCATATGTCACCTGATGCTGCGGGTTAATCAGCCTGCTTGTAAACTGTCAACAT 2599
Qy 2581 GGGAAACAGAGTAGGAGAGAGCTGCGAGCTGTGAGGAGTAGGAGAGGAACTTAC 2640
Db 2600 GGGAAACAGAGTAGGAGAGAGCTGCGAGCTGTGAGGAGTAGGAGAGGAACTTAC 2659
Qy 2641 GGAGTGTCCGTGCGCTGACCAACAGCATCTCTGCTATCATTTCTTTGGCAATACT 2700
Db 2660 GGAGTGTCCGTGCGCTGACCAACAGCATCTCTGCTATCATTTCTTTGGCAATACT 2719
Qy 2701 TTAATGAGTATGACCAATGCACTTTTATTTGGTGTATCATATGAAGAGGCTCTACAGG 2760
Db 2720 TTAATGAGTATGACCAATGCACTTTTATTTGGTGTATCATATGAAGAGGCTCTACAGG 2779
Qy 2761 CCACCTTAGACCAAGCACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAACTTCCAGT 2820
Db 2780 CCACCTTAGACCAAGCACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAACTTCCAGT 2839
Qy 2821 AATATTTGTCAGGACAGATTTAAACAAAGTTGCTGACCTGTGCTTTCCGCTCGGTCTGAT 2880
Db 2840 AATATTTGTCAGGACAGATTTAAACAAAGTTGCTGACCTGTGCTTTCCGCTCGGTCTGAT 2899

QY 2881 GCTGATCAGTCTGGCTGTGACCTCTCTGCTCTGCTCTTACATACCTGTTCTTAGTA 2840
Db 2900 GCTGATCAGTCTGGCTGTGACCTCTCTGCTCTGCTCTTACATACCTGTTCTTAGTA 2959
QY 2941 AATGAAGTTGGAGTCAGTTAGCTGCTATGACATGCTGTATGCTGCGAGACCTACTGGGA 3000
Db 2960 AATGAAGTTGGAGTCAGTTAGCTGCTATGACATGCTGTATGCTGCGAGACCTACTGGGA 3019
QY 3001 TTGGATAAATTTAGGCTCTCCCTTTCTGGAGATGCTGCGCCGAGAGATGGCAAGATCGATGC 3060
Db 3020 TTGGATAAATTTAGGCTCTCCCTTTCTGGAGATGCTGCGCCGAGAGATGGCAAGATCGATGC 3079
QY 3061 TTGGAGTGTAGAGAGCCGACAGGCTCTCTGCGGAACTGTAGAGAAATTTGAGCAG 3120
Db 3080 TTGGAGTGTAGAGAGCCGCGAGGCTCTCTGCGGAACTGTAGAGAAATTTGAGCAG 3139
QY 3121 GCAGGAGGAGGAGGACATTTGATGCTGCGGCTCTTACTTACTCTAGTACATAGACCCAC 3180
Db 3140 GCAGGAGGAGGAGGACATTTGATGCTGCGGCTCTTACTTACTCTAGTACATAGACCCAC 3199
QY 3181 GTCATATCAGCTGGAGTCAATCAGAGGCGCGAGACTATCACAGGCTCTGATGCC 3240
Db 3200 GTCATATCAGCTGGAGTCAATCAGAGGCGCGAGACTATCACAGGCTCTGATGCC 3259
QY 3241 TCAGGCGCTGAAGCAAAAGTCCAGGAGAGAGCATGACCTTTGACGATGATCACC 3300
Db 3260 TCAGGCGCTGAAGCAAAAGTCCAGGAGAGAGCATGACCTTTGACGATGATCACC 3319
QY 3301 ACTGGTGTCTTATCAAGTGTCCCAAAATGAAAAATTTCTACATCTTACGAGGAAAGA 3360
Db 3320 ACTGGTGTCTTATCAAGTGTCCCAAAATGAAAAATTTCTACATCTTACGAGGAAAGA 3379
QY 3361 CGGAAGCAAGCTACCGCTATTGTTTACTTGGAGTAATAGGAGCTGAAATTTGGTGTGAA 3420
Db 3380 CGGAAGCAAGCTACCGCTATTGTTTACTTGGAGTAATAGGAGCTGAAATTTGGTGTGAA 3439
QY 3421 ATTGAACCTCTAAACTATTGACAGAGCTCGAAGCTCTAGCCAAATTCCTGAGGATTC 3480
Db 3440 ATTGAACCTCTAAACTATTGACAGAGCTCGAAGCTCTAGCCAAATTCCTGAGGATTC 3499
QY 3481 GGGTTGACTAGTGTGTGATCAACTACTGCTGCGCAGACATCTTGCAAGGCACTGACG 3540
Db 3500 GGGTTGACTAGTGTGTGATCAACTACTGCTGCGCAGACATCTTGCAAGGCACTGACG 3559
QY 3541 TTTCTTTCTGTCAGAGCTCCAGCCCAAACTTCTCCACACAGCACTATCCGAAGAACA 3600
Db 3560 TTTCTTTCTGTCAGAGCTCCAGCCCAAACTTCTCCACACAGCACTATCCGAAGAACA 3619
QY 3601 GCCATTGATCTGATTTGGAAGTGGGTTCACTGTTGGAGGCTTACATGAGATGTCCGCT 3660
Db 3620 GCCATTGATCTGATTTGGAAGTGGGTTCACTGTTGGAGGCTTACATGAGATGTCCGCT 3679
QY 3661 GTTCTGATGGGCTTCTCGAACTTTGCGGATGCGGAGAAACAATTCGCCACATCACA 3720
Db 3680 GTTCTGATGGGCTTCTCGAACTTTGCGGATGCGGAGAAACAATTCGCCACATCACA 3739
QY 3721 ATGGGGTTGCTCTGAGCCAGCAGCTGACTCGGCGCGCTCTGCGAGGATGCCCTCTCG 3780
Db 3740 ATGGGGTTGCTCTGAGCCAGCAGCTGACTCGGCGCGCTCTGCGAGGATGCCCTCTCG 3799
QY 3781 CTCATTGGCACCGCCAGACACCCGCTTTCATCACACATAGCCAAAGAGGTACACAGA 3840
Db 3800 CTCATTGGCACCGCCAGACACCCGCTTTCATCACACATAGCCAAAGAGGTACACAGA 3859
QY 3841 CATACGGCTTTGACAGAAATACCCCAATCAGAGCAATATGCACACAACTCTTGA 3900
Db 3860 CATACGGCTTTGACAGAAATACCCCAATCAGAGCAATATGCACACAACTCTTGA 3919
QY 3901 CGAGCTAAAGGGGAAATTTGAGAGTCATTGAAATTTCTTATTGAAAAGATGCCACAGAT 3960
Db 3920 CGAGCTAAAGGGGAAATTTGAGAGTCATTGAAATTTCTTATTGAAAAGATGCCACAGAT 3979
QY 3961 GTTGTGGATTTCTCTGTTGGAGGTTATGGACATCATTTATGTACTGCTTGAAGGATCTTTA 4020

Db 3980 GTTGTGGATTTCTCTGTTGGAGGTTATGGACATCATTTATGTACTGCTTGAAGGATCTTTA 4039
QY 4021 GTTAAAAAGAAAGGCTCTTCAAGAAATGTTTCCAGCATCTGCGAGGTTCTACATGGTCAGC 4080
Db 4040 GTTAAAAAGAAAGGCTCTTCAAGAAATGTTTCCAGCATCTGCGAGGTTCTACATGGTCAGC 4099
QY 4081 TATTATGAGCGGAATCACAAGATAGCAGTGTGGAGCTGCGCATCTGAGTGGCCCTGTAC 4140
Db 4100 TATTATGAGCGGAATCACAAGATAGCAGTGTGGAGCTGCGCATCTGAGTGGCCCTGTAC 4159
QY 4141 GACATCCGAGCTGGAATGTCAGACATTCATGGGACACAGGACCAATCAGTGCAGTG 4200
Db 4160 GACATCCGAGCTGGAATGTCAGACATTCATGGGACACAGGACCAATCAGTGCAGTG 4219
QY 4201 GCTTTTGTCTCTGATGGAAGATATCTTGCCACCTACTCAAACTGACAGCCACATTTCT 4260
Db 4220 GCTTTTGTCTCTGATGGAGATATCTTGCCACCTACTCAAACTGACAGCCACATTTCT 4279
QY 4261 TTTTGGCAGATGAACACGTCAGCTGTGGGAAGCATCGGCAATGTGAATCGGCACCTCAG 4320
Db 4280 TTTTGGCAGATGAACACGTCAGCTGTGGGAAGCATCGGCAATGTGAATCGGCACCTCAG 4339
QY 4321 CTGCGCTGATTTAAACCTTACAGGTGCCCCCTGTGAGCCGCGCTCCCCGGCTCCAC 4380
Db 4340 CTGCGCTGATTTAAACCTTACAGGTGCCCCCTGTGAGCCGCGCTCCCCGGCTCCAC 4399
QY 4381 AATGCCCTCAGCTGCGCGCTCATCTGAGCTTCCAAACCGCAACGTCATCTCATGGCC 4440
Db 4400 AATGCCCTCAGCTGCGCGCTCATCTGAGCTTCCAAACCGCAACGTCATCTCATGGCC 4459
QY 4441 CATGACGGGAAGGAGCAGCCGCTTCATGTCCTAA 4473
Db 4460 CATGACGGGAAGGAGCAGCCGCTTCATGTCCTAA 4492

RESULT 4
AAD18131
ID AAD18131 standard; cdna; 4497 BP.
AC AAD18131;
XX 18-DEC-2001 (first entry)
DT Mouse GF-beta Resistance Associated Gene (TRAG) cdna.
DE Mouse; Transforming growth factor-beta; TGF-beta; cellular proliferation;
KW phenotype; TRAG; TGF-beta Resistance Associated Gene; cancer therapy;
KW oncogenesis; gene mapping; transgenic animal; chromosome 18; ss.
XX Mus sp.
XX Key Location/Qualifiers
FT CDS 28..4497
FT /*tag= a
FT /product= "Mouse TRAG protein"
XX WO200166739-A1.
XX 13-SEP-2001.
XX 12-FEB-2001; 2001WO-US004475.
XX 07-MAR-2000; 2000US-0187572P.
XX 08-MAR-2000; 2000US-0187848P.
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX Thorgerirsson SS, Sanders S;
XX WPI; 2001-589942/66.
XX P-PSDB; AAE10803.
XX

PT Novel isolated polypeptide for studying and modulating mechanisms
PT involved in cellular proliferation comprises transforming growth factor-
PT beta resistance associated polypeptide fragment.

The present sequence is a cDNA encoding mouse transforming growth factor (TGF)-beta resistance associated (TRAG) protein. Mouse TRAG gene is located on chromosome 18. TRAG DNA or protein is useful for studying and modulating mechanisms involved in cellular proliferation, and for modulating cellular phenotype. It is also used for screening aggressive metastasis in cancer cells, and for targets used in cancer therapy. TRAG protein is useful for evaluating factors that interact with and/or control TGF-beta signalling for understanding both cell proliferation control and oncogenesis, and in protein-protein interaction cells. TRAG DNA is useful as hybridisation probe, in chromosome and gene mapping, in the generation of antisense RNA and DNA, for the preparation of TRAG protein by recombinant techniques, and for generating either transgenic or knockout animals which, in turn, are useful in the development and screening of therapeutically useful reagents

Sequence 4497 BP: 1144 A; 1153 C; 1147 G; 1053 T; 0 U; 0 Other;

Qy	1	ATGGCAGAAACAGCCTTGTGTTCTACCCATTTGTTCTTTGGGGTGCAGAAAGCGCCACACAT	60
Db	28	ATGGCAGAAACAGCCTAGTTCTTGCCCATTTGTTCTTTGGGGCCGCAAGACACCCACACAT	87
Qy	61	TGCATCTCAGCGGTACTTTTAAACAGATCATGGGGCCAGATCGTTAAACAGGATGTCAACGAC	120
Db	88	TGCATTTTCGTCAATACTGTTGACAGATGATGGGGCACAATTTGTAATCTGGATGCCACGAT	147
Qy	121	GGACAAATATGTCTCTGGGATCTTTTCAGTAGAAATGCAAAATTAATCTCTCGAGCACTGTTG	180
Db	148	GGACAAATATGTCTCTGGGATGTTTCGTGTAGAACTAGAAAGTTAATCCCGAGCACTGTTA	207
Qy	181	TTTGGTCAATACAGCATCAATCACTTGTGTTTGTCTAAAGCTTTGTGCTTCCAGTGCACAAACAG	240
Db	208	TTTGGCCACACAGCATCCATCACTTGTGTTTGTCAAAAGCCTGCGCTTCTTGGGGACAAGCGG	267
Qy	241	TATATTGTGAGTGCAATCTGAAAGTGGACAGATGTGCCCTCTGGGATGTCAAGTATGGCAGA	300
Db	268	TACACTGTGAGCGCGCTCTGCAACCGAGAGATGTGCCCTCTGGGATGTCAACGATGGCAGA	327
Qy	301	TGTATTGGAATTTACAAATATTAGCTTTGCACACATCTGGCATAACAGTTCTTACCAAGTTCTCT	360
Db	328	TGTATTGGAATTTACCAAGTTAGCGCTGCACACACACTGGCATAACAGTTCTTACCAAGTTCTCT	387
Qy	361	GTTGGGAATCAGCGAGAAAGGAGGCTTTTATAGCCACGACCAATTAACCTTGAAATCCTTGTT	420
Db	388	GTTGGGAACCAAGCAAGAGGCGAGGCTCTCTGCCATGGACATTTAGCCTGAAATCCTCGTT	447
Qy	421	GTGGATGTACAGCGCTTGAAGTATTATACCTCTTAGTATCAAGATATCAACGACATGG	480
Db	448	GTGGATGCCACAGCGCTTGAAGTGTGTATTCTCTGGTATCGAAGATCTCTCCAGACTGG	507
Qy	481	ATTAGCTCCATCAGTATTATTTCGATCCCAACCGAAACAAGAGGACAAGTGTGACATCTC	540
Db	508	ATCAGTCCATGAGCATCATCCACTCTCAGCGGACACAAGAGGACACTGTGTGGCGCTC	567
Qy	541	TCGGTGAATGGCATCCTGAAGGCTGTGGATTGTTAACCTCGGAAATGAAGTGCACATGACAGAT	600
Db	568	TCTGTGACAGGTATTCTGAAGGTGTGGATTGTCACTCTGAAATGAGTGGAAATGAGGAT	627
Qy	601	ACTGAGCCCAATTTTGGAGGGAATCCAAACCAATTTATGTTCAGAAATGGCCAAAGCATC	660
Db	628	ACTGAGCCCAATTTTGGAGGAGGATCCAAACCAATTTATGTTCAGAAATGGCCAAAGCATC	687
Qy	661	TCTTTTGTGTGCATTTTACAAAGGTCACTTTTGGTGTGTGTTCCTCAAAATATTGGAGGGTG	720

1801 ATACAGCAGTTGAGATTTCTAAAGCTTTGATGAGAGCTGTTCTGCTGCTGTTGATTTCA 1860
1828 ATACAGCAGTTGAGATTTCTAAAGCTTTGATGAGAGCTGTTCTGCTGCTGTTGATTTCA 1887
1861 CTTAGTCATCCAGCAGTCAACCTTAAACAGCTTATGACGAGCTAGTCTTCTGCTGCTT 1920
1888 CTTAGTCATCCAGCAGTCAACCTTAAACAGCTTATGACGAGCTAGTCTTCTGCTGCTT 1947
1921 AAAAATATGCCCCATCATAGCTACAAACCTTTGCAACTAACTCTTTGGCTTCTGAGGCA 1980
1948 AAAAATATGCCCCATCATAGCTACAAACCTTTGCAACTAACTCTTTGGCTTCTGAGGCC 2007
1981 TCTGCAAGGGAATTTACTTAATATTTCTATAACTCTCTGATGCTTCTGATGCTTCTGATG 2040
2008 TCTGCAAGGGAATTTACTTAATATTTCCATAACTCTCTGATGCTTCTGATGCTTCTGATG 2067
2041 ACAAACTACAGCCCGGACATACATGCTGCTTCTTTGATGCTGAGGCTTGTATTT 2100
2068 ACAAACTACAGCCCGGATATCCATGCTTCTTTGATGCTGAGGCTTGTATTT 2127
2101 CAATCTCTGACTGAAGAGCTCTAGGCCGAATACTGCTCTTATTTTCCCGCAGAGATTTG 2160
2128 CAATCTCTGACTGAGAGGCTCTAGGCCGAATACTGCACTTATTTTCCCGCAGAGATCTG 2187
2161 CAAAAAGCATCTGGCAGTTTCTGAGCAAAAGGGGCTCTTTTAACTGGAAGGAGGAGCA 2220
2188 CAGAAAGCATCTGGCAGTTTCTGAGCAAAAGGGGCTCTTTTAACTGGAAGGAGGAGCA 2247
2221 GTTCTCTTCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2280
2248 GTTCTCTTCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2307
2281 GATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2340
2308 GAG 2367
2341 CGGTCCAGCAAAATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2400
2368 CGGGCCAGCAAGTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2427
2401 AAGCTGTTATGCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2460
2428 AATATTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2487
2461 CTGAGTCCGCTTGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2520
2488 CTGAGTCCGCTTGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2547
2521 GGAGGCCATATGTCAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2580
2548 GGAGGTCATATGTCGCTGATGCTTCCCGGTTAATATCAGGCTGCTGGAAGGAGGAGGAG 2607
2581 GGGAAACAGAGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2640
2608 GCGAAAGCAGAGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2667
2641 GAGAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2700
2668 ACAGTGTCTGAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2727
2701 TTAATGAGTATGACCAATGCAATTTTATTTGATGATCATATGAGGAGGAGGAGGAGGAGGAG 2760
2728 TTAATGAGTATGACCAATGCAATTTTATTTGATGATCATATGAGGAGGAGGAGGAGGAGGAG 2787
2761 CCACTTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2820
2788 CCGCTTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2847
2821 AATATTGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2880
2848 AACATTGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2907

2881 GCTGATCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2940
2908 GCTGATCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2964
2941 AATGAAGGTTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3000
2965 AATGAAGGTTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3024
3001 TTGGATAAATTTAGGCTTCCCTTCTGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3060
3025 CTGAGAGATTCAGGCTTCTCTCTCTGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3084
3061 TTGAGGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3120
3085 TTGAGGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3144
3121 GCAGGAG 3180
3145 GCAGGAG 3204
3181 GTCATATCACCTGAGTCTACATCAGAAAGCGGCGAGAGTATCACACCGGCTCTCTGATGCC 3240
3205 GTCATATCACCTGAGTCTACATCAGAAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3264
3241 TCAGGGCTTGAAGCAAAAGTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3300
3265 TCAGGGCTTGAAGCAAAAGTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3324
3301 ACTGGTGTCTTATCAAGTGTCCCAAAAGTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3360
3325 GCAGGTTGTCTGAGTGTCCCAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3384
3361 CGGAAGCAAGTACCGCTTATTTTATCTTGAAGTATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3420
3385 AGGAG 3444
3421 ATTGAACCTCTTAACATATGACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3480
3445 ATTGAACCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3504
3481 GGGTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3540
3505 GGGTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3564
3541 TTTCTTCTGCTACAGGCTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3600
3565 TATCTTCTGCTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3624
3601 GCAATGATCTGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3660
3625 GCACTGAGTCTGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3684
3661 GTTCTGATGGGCTTCTGAACTTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3720
3685 GTTCTGATGGGCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3744
3721 ATGGGCTTCCCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3780
3745 ATGGGCTTCCCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3804
3781 CTCATTTGAG 3840
3805 CTCATCGGAG 3864
3841 CATATGAG 3900
3865 CACAGGAG 3924
3901 CGAGCTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3960
3925 AGGGCTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3984
3961 GTTGTGAGTCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4020

Db 688 TCTTTTGTGCAATTCACAGAGGTGCTCTTGTGTTGTAATGCTCAAGTACTGGAGGGTG 747
Qy 721 TTGATGCGGGAGACTATTCCTTGTGTGTTTCAAGTCTAGTGAATAAGACAGACATGG 780
Db 748 TTGATGCTGGCGACTACTCCCTGCTGTGTTTCAAGTCTAGTGAAGATGACAGACATGG 807
Qy 781 ACCGGGGGAGCTTTGTCTCATCAGATAAAGTCAATTTGGACAGAAATGGCAAGT 840
Db 808 ACTGGAGGGGACTTTGTGTCTGACAGACAAAGTCAATTTGGACTGAATAACGGGCGAGT 867
Qy 841 TATATTTACAACTACCTGCGAGTTGCTTCCAGCTAGTGAATTCATTTCCGAGTGTGTTG 900
Db 868 TACATTTACAACTCCCTGCGAGTTGCTTCCAGCTAGTGAATTCATTTCCGAGTGTGTTG 927
Qy 901 GGGAGGCGAGTTGAAATTTAAATTCCTCTGTGTACAAATATTCCTTTGGATCGAAAGAT 960
Db 928 GGGAAAGCAGTTGAAATCTGATTCCTCCGTCGAGCATAGACCTCTTGGATCAGAGGAT 987
Qy 961 AAGAGTTGCTAATTTGCTCTCTGTTACTCTGTTTCTTCTATGATGCGAGAAATTTTC 1020
Db 988 AGAGAGTTGTAATTTGCTCTCTGTTACTCTGTTTCTTCTATGATGCGAGAAATTTTC 1047
Qy 1021 CATAAAGCTGTAATTTCAAGGTGATTTCTTGGAGGTTGAATTTTGGAAACATATCAGAC 1080
Db 1048 CATAGCTACTAATTTCAAGGTGATTTCTTGGAGGTTAAGTATTTGGAAACATAGCAGAC 1107
Qy 1081 ACAGCTGATAAACAAGGAGTGAAGAGGCTGGCAATGATCAACTTCTATTAGTTTGCAG 1140
Db 1108 ATAGCAGACAAACAGGAAGCCAAATGAAGGCTTAAACGACAACTTGTATTAGTTTGCAG 1167
Qy 1141 GAGGCATTTGATAAATCTGATTCCTGCTGCTGCTGGAATTTATAGATCAGCTGAGTGTGAT 1200
Db 1168 GATGCAATTTGACAACTGAAGCCCTGCTGCTGGAATTTATCGATCAGCTGAGTGTGAT 1227
Qy 1201 CCCAATAGTAATGAACTCTTAAAGTAATCGCAAGTGTGTACATACAGCACATGACGGA 1260
Db 1228 CCAACAGCAACGAAACCACTTAAAGTAACTTGGAGTGTGTACATACAGCACACGGCGC 1287
Qy 1261 CTGTGTTTGTGCTGGAAGTGAAGCATAGTTATTGTACTCTGCCACACAGACGGGCATA 1320
Db 1288 CTGTGTTTGGCGCGGGAAGACGAGCATCAATATCGTCCCTGCGCACCCAGACGGCCATA 1347
Qy 1321 GTACAGCTGTTGCAAGGGGAACACATGCTCAGAAAGGTTGGCCACCTCAGACAACTC 1380
Db 1348 GTTCAGCTGCTGCGGAGAACACATGCTCAGACGAGTTGGCCCGCCGCGACAGACCCCTC 1407
Qy 1381 CGTGGTCACTCGGAACAAAGTCAATGTTTCTATATCTCATCAGTCTCAGCTCGGTAT 1440
Db 1408 CGTGGCCACCGGAACAAAGTCAATGTTTCTGTATCTCATCAGGTCTCAGCTCGGTAT 1467
Qy 1441 GATCAAGATACCTGATATCTGGAGGTGTGGATTTTTCAGTCAATAATTTGGGACATATTT 1500
Db 1468 GATCAAGATACCTGATATCCGAGGTGTGGATTTTTCGTCATCATTTTGGGACATTTT 1527
Qy 1501 TCTGGAGAAATGAAACATATCTTCTGTGTTTCAATGGTGTGAGATTAATCAACTTCTAGTT 1560
Db 1528 TCCGGAGAAATGAAACATATCTTCTGTGTTTCAATGGTGTGAGATTAATCAACTTCTAGTT 1587
Qy 1561 CCACCTGAAACCTGTAGTGAAGTACAGCATGCTGCTGTGCTGAGCCAGTACCCAC 1620
Db 1588 CCGCCAGAAACCTGTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1647
Qy 1621 TCAGTAGGACTTCTAAGTTTGGAGAGAAATATGATATGTTGGCATCTCTGTCACCTT 1680
Db 1648 TCTGTAGGGCTGTAACTGTGCGAGAGAAATATGATATGATGTTGGGCTCTCTGACCTG 1707
Qy 1681 TTTCTTATCAAGTAATCAAAATGGAGCCCTTCTGATGATTAATCTGTTGGTGGGGTGTCA 1740
Db 1708 TTCCCTATTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1767
Qy 1741 GATGGTCTGTGCTGCGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1800
Db 1768 GACGGCTCTGTGTGTGCGCAGATGGAACATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1827

Qy 1801 ATAAACAGAGTTCAGATTTCTAAACGCTTGTGATGAAGCTGTTCTGCTGCTGTGATTC 1860
Db 1828 ATAAACAGCGGTGAGATTTCTCAATGCTTGTGACGAAGCTGTCCTGTCAGCAGTGGACTCA 1887
Qy 1861 CTTAGTCTATCCAGCAGTCAACCTTAAACAGCTATGACGAGAGTGTCTTCTGCTCTT 1920
Db 1888 CTGAGTCAACCCAGCAGTCAACCTTAAACAGCTATGACGAGGAGTGTCTGCGCGCTT 1947
Qy 1921 AAAAATATGCCCCATCATAGCTACAAACCTTTGCAACTAACTCTTTGGCTTCTGAGGCA 1980
Db 1948 AAAAATGCCCCACCAACAGCTGCAACCTTTGCAACTAACTCTTTGGCTTCTGAGGCC 2007
Qy 1981 TCTGCAAGGGAAATTTTACCTAAATATCTCATAACTCCCTGATGTTTCAAGCAATAAAG 2040
Db 2008 TCTGCAAGGGAAATTTTACCTAAATATCTCATAACTCCCTGATGTTTCAAGCAATAAAG 2067
Qy 2041 ACAAACCTTAAACAGCCGACATACATGCTGCTATTTCTTGTATGATGGAAGGTTGATTT 2100
Db 2068 ACAAACCTTAACTGACCCGATATCCATGCTCTTTCTTGTATGGAAGCTTTGATTTAT 2127
Qy 2101 CAACTCTGACTGAAGAGCCTCTAGGCCGAATACTGCTTATTTTCCCGCAGAGAAATTTG 2160
Db 2128 CAACTCTGACTGAAGAGCCTCTAGGCCGAATACTGCTTATTTTCCCGCAGAGAAATCTG 2187
Qy 2161 CAAAAAGCATCTGGCAGTTTCAAGAAAGGGGCTCTTTTAACTGGAAGAACAGCAGCA 2220
Db 2188 CAGAAAGCATCTGGCAGTTTCAAGAAAGGGGCTCTTTTCTGACTGGAAGAACAGCAGCA 2247
Qy 2221 GTTCTCTTCCAAAGTGAAGAGCATCAAGAGAACTCAAGAAACACTCTCTTGTAT 2280
Db 2248 GTTCTTTTCCAGCAAGTGAAGAACTATCAAGAGAACTATCAAGAGAACTCTCTTGTAT 2307
Qy 2281 GATGAAGAGGAGTCAAGAGATTAATGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2340
Db 2308 GAGGAGAGGAGCAGGAGAGGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2367
Qy 2341 CGTCCAGCAATCAAGGCTTGTACCTTATAGAAATATAATTTAACTATGACACTGCA 2400
Db 2368 CGGGCAGCAAGTCCAAACCACTTACCTTACTAGATAACACCTTACTATGATACCGCA 2427
Qy 2401 AAGCTGTTTATGCTCTGCTTCAAGCTGCGGTTTGAAGTACTGAGTGAAGTTTGC 2460
Db 2428 AAATATTATGCTCTGCTTCAAGCTGCGGTTTGAATGAAGTTTCTGAGTGAAGTTTGC 2487
Qy 2461 CTGATGCTGCTGGAATGCTGAAACCCCACTGACCGTATCGTTTGGCTCTTGTCAAGA 2520
Db 2488 CTGATGCTGCTGGAATGCTGAAACCCCACTGACCGTATCGTTTGGCTCTTGTCAAGA 2547
Qy 2521 GGAGGCTATGCTGCTGATGCTGCGGTTTAAATCAGCCTGCTTGTAACTGTGACAT 2580
Db 2548 GGAGGCTATGCTGCTGATGCTTCTGTTTAAATCAGCCTGCTGGAAGCTTACTGAC 2607
Qy 2581 GGGAAAACAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2640
Db 2608 GCGAAAGCAGAGCAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2667
Qy 2641 GGAGTGTCCGCTGCGCTGACACAGCATCTCTGCTTATCATTTTCTTGGCAATACT 2700
Db 2668 ACAGTGTCCGAGCGGTGACCGCAACATCTGTTGTCTCATATATCTCTGGAATACT 2727
Qy 2701 TTAATGAGTATGACCAATGCAACTTTTATTTGTTGATCATATGAAGAGGAGGCTTAC 2760
Db 2728 TTAATGAGTATGACCAATGCGAGTTTCAATTTGAGATCATATGAAGAGGAGGAGGAGGAG 2787
Qy 2761 CCACTTAGACCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2820
Db 2788 CCGCTTAGACCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2847
Qy 2821 AATATTGTGCAAGGAGCAGATTAACAGATGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 2880
Db 2848 AATATTGTGCAAGGAGCAGATTAACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2904

1677 CCATATGTCACTGCTGCGGGTTATATCAAGCTGCTGTAAACTGTCACATCGGAA 1736
2586 AACAGAAGTAGGAAGAGTGCAGCGTCTGAGGAGTAGGAAGAGGAACCTTACCGAGT 2645
1737 AACAGAAGTAGGAAGAGTGCAGCGTCTGAGGAGTAGGAAGAGGAACCTTACCGAGT 1796
2646 GTCCCGTGGCGTCAACACAGCATCTCTGCTATCATCTTCTTCCGCAAACTATTAAT 2705
1797 GTCCCGTGGCGTCAACACAGCATCTCTGCTATCATCTTCTTCCGCAAACTATTAAT 1856
2706 GAGTATGACCAATGCAACTTTTATGGTGATCATATGAAGAAGGGTCTTACCAGGCCACC 2765
1857 GAGTATGACCAATGCAACTTTTATGGTGATCATATGAAGAAGGGTCTTACCAGGCCACC 1916
2766 TAGACCAAGCACCCAGACCTTTCTAAGGCAAGGGTTCCTCCCAACTTCCAGTAAATAT 2825
1917 TAGACCAAGCACCCAGACCTTTCTAAGGCAAGGGTTCCTCCCAACTTCCAGTAAATAT 1976
2826 TGTGCAAGGACAGATTAAACAGTTGCTGCACCTGCTGTTCGCTCGGTCTGATCTGA 2885
1977 TGTGCAAGGACAGATTAAACAGTTGCTGCACCTGCTGTTCGCTCGGTCTGATCTGA 2036
2886 TCACTCTGGCTCTGACCCCTCTTCTGCTCTGCTCTTACATACCTGTTCTTAGTAAATGA 2945
2037 TCACTCTGGCTCTGACCCCTCTTCTGCTCTGCTCTTACATACCTGTTCTTAGTAAATGA 2096
2946 AGGTGGAGTCAGTATGCTATGCACTGTGTATGCTGCGCAGACTCTAGGGATTGGA 3005
2097 AGGTGGAGTCAGTATGCTATGCACTGTGTATGCTGCGCAGACTCTAGGGATTGGA 2156
3006 TAAATTTAGGCTTCCCTCTGAGATGCTGCGCCGGAAGTGGCAAGATGCGATGCTTGA 3065
2157 TAAATTTAGGCTTCCCTCTGAGATGCTGCGCCGGAAGTGGCAAGATGCGATGCTTGA 2216
3066 GGTGAGAGAAGCCGACAGGCCCTGCTTCTGCGGGAACCTGAGAAGATTGAGCAGGCAGG 3125
2217 GGTGAGAGAAGCCGACAGGCCCTGCTTCTGCGGGAACCTGAGAAGATTGAGCAGGCAGG 2276
3126 CAGGAAGGAAGCAATTGATGCTGGCTCCTTACTTACTCTCAGTACATGACCACTCAT 3185
2277 CAGGAAGGAAGCAATTGATGCTGGCTCCTTACTTACTCTCAGTACATGACCACTCAT 2336
3186 ATCACTGGAGTCACATCAGAAGCCGCGAGACTATCAACAGCCCTCTGATGCCCTCAGG 3245
2337 ATCACTGGAGTCACATCAGAAGCCGCGAGACTATCAACAGCCCTCTGATGCCCTCAGG 2396
3246 GCCTGAAGCAAAAGTCCAGGAGGAAGCATGACCTTGTGAGCATGACATCAACCACTGG 3305
2397 GCCTGAAGCAAAAGTCCAGGAGGAAGCATGACCTTGTGAGCATGACATCAACCACTGG 2456
3306 TTGCTTATCAAGTGTCCCAAAATGAAAAATTTCTACATCTTACAGGAAGACGGA 3365
2457 TTGCTTATCAAGTGTCCCAAAATGAAAAATTTCTACATCTTACAGGAAGACGGA 2516
3366 GCAAGCTACCGCTATTGTTTTTACTTGGAGTAATAGGAGCTGAATTTGGTCTCAAAATGA 3425
2517 GCAAGCTACCGCTATTGTTTTTACTTGGAGTAATAGGAGCTGAATTTGGTCTCAAAATGA 2576
3426 ACCTCTTAAACTATTGACAGACCTCGAAGCTCTAGCCAAATTCCTGAGGGATTCCGGTT 3485
2577 ACCTCTTAAACTATTGACAGACCTCGAAGCTCTAGCCAAATTCCTGAGGGATTCCGGTT 2636
3486 GACTAGTGTGGATCCAACTACTCGTGGCCAGACATATTTGCAAGGCACTGAGCTTCT 3545
2637 GACTAGTGTGGATCCAACTACTCGTGGCCAGACATATTTGCAAGGCACTGAGCTTCT 2696
3546 TCTGCTACAGCTCCCAAGCCCAAACTTCTCCACACAGCACTATCCGAGAACAGCCAT 3605
2697 TCTGCTACAGCTCCCAAGCCCAAACTTCTCCACACAGCACTATCCGAGAACAGCCAT 2756
3606 TGATCTGATTGGAGCTGGGTTCACTGTTTGGGAGCCCTTACATGGATGTGTCCGCTGTCT 3665

2757 TGATCTGATTGACGCTGGGTTCACTGTTTTGGGAGCCTTACATGATGTGTCCGCTGTCT 2816
3666 GATGGGCTTCTCGAACTTTGTGCGCATGCGCAGAAAACAACTTGCCAAATCACAAATGG 3725
2817 GATGGGCTTCTCGAACTTTGTGCGCATGCGCAGAAAACAACTTGCCAAATCACAAATGG 2876
3726 GTTGGCTCTGAGCCAGCAGCTGAGCTCGGCCCGCTCTGCGAGGCAATGCCCTCTCGCTCAT 3785
2877 GTTGGCTCTGAGCCAGCAGCTGAGCTCGGCCCGCTCTGCGAGGCAATGCCCTCTCGCTCAT 2936
3786 TGCACCGCAGACACCCCGCTTCAATCAACCAATAGCAGCAAGAGGTAACAGACATAC 3845
2937 TGCACCGCAGACACCCCGCTTCAATCAACCAATAGCAGCAAGAGGTAACAGACATAC 2996
3846 GGCTCTTGCAGCAAAATACCCCAATCACAGCAGAAATATGCAACAACTCTTGCGAGC 3905
2997 GGCTCTTGCAGCAAAATACCCCAATCACAGCAGAAATATGCAACAACTCTTGCGAGC 3056
3906 TAAAGGGGAAATTTTGGAGTCAATTTGAATTTCTTATTTGAAAAGATGCCCAAGATTTGT 3965
3057 TAAAGGGGAAATTTTGGAGTCAATTTGAATTTCTTATTTGAAAAGATGCCCAAGATTTGT 3116
3966 GGATCTTCTGCGAGGTTATGCAATCATATATGTAATGTAATGTAATGTAATGTAATGTAAT 4025
3117 GGATCTTCTGCGAGGTTATGCAATCATATATGTAATGTAATGTAATGTAATGTAATGTAAT 3176
4026 AAAGAAAGTCTTCAAGATGTTTCCAGCCATCTGCGAGTCTTACATGTCAGCTTATTA 4085
3177 AAAGAAAGTCTTCAAGATGTTTCCAGCCATCTGCGAGTCTTACATGTCAGCTTATTA 3236
4086 TGACGGGAATCACAGAATAGCAGTTGGAGCTCGCCATGTTTCAAGTGGCCCTGTGACACAT 4145
3237 TGACGGGAATCACAGAATAGCAGTTGGAGCTCGCCATGTTTCAAGTGGCCCTGTGACACAT 3296
4146 CCGGACTGGAATGTCAGACAAATCCATGGAACAAGGACCAATCACTGCGAGTGGCTTT 4205
3297 CCGGACTGGAATGTCAGACAAATCCATGGAACAAGGACCAATCACTGCGAGTGGCTTT 3356
4206 TGCTCTGATGGAAGATATCTTGGCACTTACCACTGACAGCACTGAGCACTTCTTTTGG 4265
3357 TGCTCTGATGGAAGATATCTTGGCACTTACCACTGACAGCACTGAGCACTTCTTTTGG 3416
4266 GCAGATGAACACGTCTGCTGGAAGCATCGGCACTGTAACCTCGGCACTCGAGTGG 4325
3417 GCAGATGAACACGTCTGCTGGAAGCATCGGCACTGTAACCTCGGCACTCGAGTGG 3476
4326 CTGCATTAAAACTTACAGGCTGCTGAGCCCGCTGAGCCCGCTGAGCCCGCTGAGCCCGCT 4385
3477 CTGCATTAAAACTTACAGGCTGCTGAGCCCGCTGAGCCCGCTGAGCCCGCTGAGCCCGCT 3536
4386 CCTCAAGCT-GGCCCGCTCATCTGGACTTCCAAACCGCAACGTCATCTCATGCGCCCATG 4444
3537 CCTCAAGCTGGGCGCGCTCATCTGGACTTCCAAACCGCAACGTCATCTCATGCGCCCATG 3596
4445 ACGGGAAGGAGCACCCCTTTCATGGTCTAA 4473
3597 ACGGGAAGGAGCACCCCTTTCATGGTCTAA 3625

RESULT 7

ADA53749

ID ADA53749 standard; cDNA; 3316 BP.

XX ADA53749;

XX 20-NOV-2003 (first entry)

XX Human coding sequence, SEQ ID 1317.

XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;

XX Gene Therapy; human; secretory protein; membrane proteins; cancer;

XX inflammatory disease; osteoporosis; neurological disease; gene; ss.

QY	1621	TCAGTAGGACTTCTAAGCTTTGCGAGAGAAAAATGCAATAATGTTGGCATCTCGTCACCTT	1680
DB	1713	TCAGTAGGACTTCTAAGCTTTGCGAGAGAAAAATGCAATAATGTTGGCATCTCGTCACCTT	1772
QY	1681	TTTCCTATTCAGTAAATCAAATGAGGCCCTTCGTAGATGATTAACCTGTFGTGGGGTGTTCA	1740
DB	1773	TTTCCTATTCAGTAAATCAAATGAGGCCCTTCGTAGATGATTAACCTGTFGTGGGGTGTTCA	1832
QY	1741	GATGGTTCTGTGTAAGCTCTGGCAAAATGGATACCTGGTGCATTTGGATCGTTGTGTGATGGGG	1800
DB	1833	GATGGTTCTGTGTAAGCTCTGGCAAAATGGATACCTGGTGCATTTGGATCGTTGTGTGATGGGG	1892
QY	1801	ATAACAGCAGTTGAGATTCTTAAACGCCTTGTGATGAAGCTGCTTCTGCTGCTGCTGTGATTTCA	1860
DB	1893	ATAACAGCAGTTGAGATTCTTAAACGCCTTGTGATGAAGCTGCTTCTGCTGCTGCTGTGATTTCA	1952
QY	1861	CTTAGTCAATCCAGCAGTCAACCTAAAAACAAGCTATGACGAGACGTAGTCTTGTGCTCTTT	1920
DB	1953	CTTAGTCAATCCAGCAGTCAACCTAAAAACAAGCTATGACGAGACGTAGTCTTGTGCTCTTT	2012
QY	1921	AAAAATATGGCCCATATAGCTACAAACCCCTTGCRACTAACTCTTGGCTTCTGAGGCA	1980
DB	2013	AAAAATATGGCCCATATAGCTACAAACCCCTTGCRACTAACTCTTGGCTTCTGAGGCA	2072
QY	1981	TCTGACAAGGGAAATTTACCTTAAATATTTCTATAAATCTCCCTGATGGTTTCAAGCAATAAAG	2040
DB	2073	TCTGACAAGGGAAATTTACCTTAAATATTTCTATAAATCTCCCTGATGGTTTCAAGCAATAAAG	2132
QY	2041	ACAAACCTTAAACAGACCCGGACATACATGTGCTATTTCTTTCATGTGGAAGCGTTGATTAAT	2100
DB	2133	ACAAACCTTAAACAGACCCGGACATACATGTGCTATTTCTTTCATGTGGAAGCGTTGATTAAT	2192
QY	2101	CAACTCTGACTGAAGAAGCCTCTAGCGCGAATACTGCTCTTAATTTTCCCCAGAGAATTTC	2160
DB	2193	CAACTCTGACTGAAGAAGCCTCTAGCGCGAATACTGCTCTTAATTTTCCCCAGAGAATTTC	2252
QY	2161	CAAAAGCATCTGGCAGTTTCAGACAAAGGGGGCTCTTTTTTAACTGGAACACGAGCAGCA	2220
DB	2253	CAAAAGCATCTGGCAGTTTCAGACAAAGGGGGCTCTTTTTTAACTGGAACACGAGCAGCA	2312
QY	2221	GTTCTCTTTCCAAAGTGAAGAAACGATCAAGAGAACATCAAGGAACACCTCCTTGAT	2280
DB	2313	GTTCTCTTTCCAAAGTGAAGAAACGATCAAGAGAACATCAAGGAACACCTCCTTGAT	2372
QY	2281	GATGAAGAGGAGATGAGGAGATTAATGAGGCAGAGAAAGGAGAAAGTGATCTGTAATAT	2340
DB	2373	GATGAAGAGGAGATGAGGAGATTAATGAGGCAGAGAAAGGAGAAAGTGATCTGTAATAT	2432
QY	2341	CGGTCCAGCAATCAAGGCCATTGACCCCTATTAGAATATAATTTAACTATGGAACACTGCA	2400
DB	2433	CGGTCCAGCAATCAAGGCCATTGACCCCTATTAGAATATAATTTAACTATGGAACACTGCA	2492
QY	2401	AAGCTGTTTATGCTCTGCTTTCAGCCCTGGGGTTTGAATGAAGTACTGGAATGAAGTTTC	2460
DB	2493	AAGCTGTTTATGCTCTGCTTTCAGCCCTGGGGTTTGAATGAAGTACTGGAATGAAGTTTC	2552
QY	2461	CTGATCGCCTTGGAAATGCTGAAACCCCACTGCACCGTATCGTTTGGCCCTCTTGTCAAGA	2520
DB	2553	CTGATCGCCTTGGAAATGCTGAAACCCCACTGCACCGTATCGTTTGGCCCTCTTGTCAAGA	2612
QY	2521	GGAGGCCATATGTCACCTGATGCTCGCGGTTATAATCAGCTCGCTGCTGAACCTGCACAT	2580
DB	2613	GGAGGCCATATGTCACCTGATGCTCGCGGTTATAATCAGCTCGCTGCTGAACCTGCACAT	2672
QY	2581	GGGAAAAACAGAAATGAGGAAGCTGTCAGCGCTCTGAGGGAGTAGGAAAGGGAACCTTAC	2640
DB	2673	GGGAAAAACAGAAATGAGGAAGCTGTCAGCGCTCTGAGGGAGTAGGAAAGGGAACCTTAC	2732
QY	2641	GGAGTGTCCGTGCGGTCAACACACAGCATCTCTGTCTATCATTTCTTTGGCAAAATPACT	2700
DB	2733	GGAGTGTCCGTGCGGTCAACACACAGCATCTCTGTCTATCATTTCTTTGGCAAAATPACT	2792

QY	2701	TTAATGATGATGACCAATGCAACTTTTATTGGTGATCATATGAAGAAGGCTCTTACCAGG	2766
DB	2793	TTAATGATGATGACCAATGCAACTTTTATTGGTGATCATATGAAGAAGGCTCTTACCAGG	2852
QY	2761	CCACTAGACCAAGCACCCAGACCTTTCTTAAGCGAAGGGGTTCCTCCCAACTTCCAGT	2820
DB	2853	CCACTAGACCAAGCACCCAGACCTTTCTTAAGCGAAGGGGTTCCTCCCAACTTCCAGT	2912
QY	2821	AATATTGTGAAGGACAGATTAAACAAGTTTGCTGCACCTGCTCGTTTCGGCTCGTCTGAT	2880
DB	2913	AATATTGTGAAGGACAGATTAAACAAGTTTGCTGCACCTGCTCGTTTCGGCTCGTCTGAT	2937
QY	2881	GCTGATCACTCTGGCTCTGACCCCTCTTCTGCTCTGCTTTACATACCTGTTCTTTAGTA	2940
DB	2938	-----	2937
QY	2941	AATGAAGTTGGAGTCAGTTAGCTGCTATGCACTGTGTTATGCTGCCAGACCTACTGGGA	3000
DB	2938	----AAGGTTGGAGTCAGTTAGCTGCTATGCACTGTGTTATGCTGCCAGACCTACTGGGA	2993
QY	3001	TTGGATAAATTTAGGCTCTCCCTTCTGGAGATGCTGGCCCGAAGATGGCAAGATCCGATGC	3060
DB	2994	TTGGATAAATTTAGGCTCTCCCTTCTGGAGATGCTGGCCCGAAGATGGCAAGATCCGATGC	3053
QY	3061	TTGAGGTGAGAGAAGCGCACAGGCGCTCTTCTGGCGGAACTGAGAGAAATTGAGCAG	3120
DB	3054	TTGAGGTGAGAGAAGCGCACAGGCGCTCTTCTGGCGGAACTGAGAGAAATTGAGCAG	3113
QY	3121	GCAGGAGGAAGGAGCCATGATGCTCTGGGCTCCTTACTTACTCAGTACATAGACCAC	3180
DB	3114	GCAGGAGGAAGGAGCCATGATGCTCTGGGCTCCTTACTTACTCAGTACATAGACCAC	3173
QY	3181	GTCATATCACTCGAGTCAATCAGAAGCGCGCAGACTATCACACGGCTCCTGATGCC	3240
DB	3174	GTCATATCACTCGAGTCAATCAGAAGCGCGCAGACTATCACACGGCTCCTGATGCC	3233
QY	3241	TCAGGGCTGAAAGCAAAAGTCCAGGAGGAGAGCATGACCTTGTGACGATGACATCACC	3300
DB	3234	TCAGGGCTGAAAGCAAAAGTCCAGGAGGAGAGCATGACCTTGTGACGATGACATCACC	3293
QY	3301	ACTGGTTGCTTATCAAGTGTCCTCC	3323
DB	3294	ACTGGTTGCTTATCAAGTGTCCTCC	3316
RESULT 8			
ADM19384			
ID ADM19384 standard; cDNA; 834 BP.			
XX	AC	ADM19384;	
XX	DT	20-MAY-2004 (first entry)	
XX	DE	Novel human channel/transporter gene #181.	
XX	KW	ds; gene; immunosuppressive; antiarthritic; antirheumatic;	
KW	KW	antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;	
KW	KW	nootropic; neuroprotective; antibacterial; virucide; fungicide;	
KW	KW	ophthalmological; gene therapy; channel/transporter protein;	
KW	KW	rheumatoid arthritis; neoplasm; cardiac arrest; cerebrovascular disorder;	
KW	KW	cerebral ischemia; angiogenesis; nervous system disorder;	
KW	KW	Alzheimer's disease; ocular disorder; corneal infection; wound healing;	
KW	KW	epithelial cell proliferation; skin aging; sunburn; transplantation;	
XX	OS	chemotaxis; food additive.	
XX	OS	Homo sapiens.	
XX	PN	WO200154472-A2.	
XX	XX	02-AUG-2001.	
PD	PF	17-JAN-2001; 2001WO-US001307.	
XX	XX		

PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-022547P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.

PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-476159/51.
XX P-PSDB; ADM19863.
XX Isolated nucleic acid molecule encoding a channel/transporter protein is
PT used in preventing, treating or ameliorating a medical condition.

XX PS Claim 1; SEQ ID NO 191; 809pp; English.

XX CC The invention relates to an isolated nucleic acid molecule encoding a

CC channel/transporter protein or sequences at least 95% identical to a

CC these. The nucleic acids and proteins encoded by them are used to

CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,

CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used

CC in diagnosing a pathological condition or susceptibility to a

CC pathologic condition. The antibodies to the proteins can also be used

CC in alleviating symptoms associated with the disorders and in diagnostic

CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays

CC (ELISA). Disorders which are diagnosed or treated include autoimmune

CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.

CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac

CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis,

CC nervous system disorders e.g. Alzheimer's disease, infections caused by

CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.

CC The polypeptides can also be used to aid wound healing and epithelial

CC cell proliferation, to prevent skin aging due to sunburn, to maintain

CC organs before transplantation, for supporting cell culture of primary

CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can

CC also be used as a food additive or preservative to increase or decrease

CC storage capabilities. This sequence corresponds to a gene of the

CC invention.

XX SQ Sequence 834 BP; 211 A; 206 C; 208 G; 209 T; 0 U; 0 Other;

Query Match 12.7%; Score 569.2; DB 5; Length 834;

Best Local Similarity 98.6%; Pred. No. 2.3e-161;

Matches 574; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGCGAGGAACAGCCTTGTCTTACCCATGTTCTTTGGGGTCCAAAGCGCCACACAT 60

DB 220 ATGCGAGGAACAGCCTTGTCTTACCCATGTTCTTTGGGGTCCAAAGCGCCACACAT 279

QY 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACCATCGTAACAGATGTACAGAC 120

DB 280 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACCATCGTAACAGATGTACAGAC 339

QY 121 GGACAAATATGTCTCGGGATCTTTTCAGTAGAATCTGCAATTAATCTCGAGCAGCTGTTG 180

DB 340 GGACAAATATGTCTCGGGATCTTTTCAGTAGAATCTGCAATTAATCTCGAGCAGCTGTTG 399

QY 181 TTTGGTCATACAGATCAATCACTGTTGTTCTTAAAGCTTGTCTTCCAGTGCAAAACAG 240

DB 400 TTTGGTCATACAGATCAATCACTGTTGTTCTTAAAGCTTGTCTTCCAGTGCAAAACAG 459

QY 241 TATATTGTGAGTGCATCTGAACTGGAGAGATGCTCTGGGATCTGAGTGATGCGACA 300

DB 460 TATATTGTGAGTGCATCTGAACTGGAGAGATGCTCTGGGATCTGAGTGATGCGACA 519

QY 301 TGTAATTGAATTTACAAATTAGCTTGACACATCTGGCATACAGTTCTTACCAGTTCTCT 360

DB 520 TGTAATTGAATTTACAAATTAGCTTGACACATCTGGCATACAGTTCTTACCAGTTCTCT 579

QY 361 GTTGGGAATCAGCGAGAGAGGGCTTTTATGCCACGACATTAACCTGAAATCTCTGTT 420

DB 580 GTTGGGAATCAGCGAGAGAGGGCTTTTATGCCACGACATTAACCTGAAATCTCTGTT 639

QY 421 GTGGATGCTACAGCCTTGAAGTATTATATCTCTTAGTATCAAGATATACCCAGACTGG 480

DB 640 GTGGATGCTACAGCCTTGAAGTATTATATCTCTTAGTATCAAGATATACCCAGACTGG 699

QY 481 ATTAGTCTCATGAGTATTATTTCGATCCACCAACACAGAGGACACAGTGTAGCACTC 540

DB 700 ATTAGTCTCATGAGTATTATTTCGATCCACCAACACAGAGGACACAGTGTAGCACTC 759

QY 541 TCGGTGATCGGCATCTGGAAGGTCTGGATTGTTTACCTCGAA 582

DB 760 TCGGTGATCGGCATCTGGAAGGTCTGGATTGTTTACCTCGAA 801

RESULT 9

ADM19354

ID ADM19354 standard; cDNA, 630 BP.

XX AC ADM19354;

XX DT 20-MAY-2004 (first entry)

XX DE Novel human channel/transporter gene #151.

XX ds; gene; immunosuppressive; antiarthritic; antirheumatic;

KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;

KW neotropic; neuroprotective; antibacterial; virucide; fungicide;

KW ophthalmological; gene therapy; channel/transporter protein;

KW rheumatoid arthritis; neoplasia; cardiac arrest; cerebrovascular disorder;

KW cerebral ischemia; angiogenesis; nervous system disorder;

KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;

KW epithelial cell proliferation; skin aging; sunburn; transplantation;

KW chemotaxis; food additive.

XX Homo sapiens.

XX WO200154472-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001307.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 14-JUL-2000; 2000US-0217496P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 18-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226686P.

PR 22-AUG-2000; 2000US-0227182P.

PR 22-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.

PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.

PR 06-SEP-2000; 2000US-0230437P.

PR 06-SEP-2000; 2000US-0230438P.

PR 08-SEP-2000; 2000US-0231242P.

PR 08-SEP-2000; 2000US-02311243P.
PR 08-SEP-2000; 2000US-02311244P.
PR 08-SEP-2000; 2000US-02311413P.
PR 08-SEP-2000; 2000US-02311414P.
PR 08-SEP-2000; 2000US-02322080P.
PR 08-SEP-2000; 2000US-02322081P.
PR 13-SEP-2000; 2000US-02311968P.
PR 14-SEP-2000; 2000US-02323977P.
PR 14-SEP-2000; 2000US-02323998P.
PR 14-SEP-2000; 2000US-02323999P.
PR 14-SEP-2000; 2000US-02324001P.
PR 14-SEP-2000; 2000US-02324010P.
PR 14-SEP-2000; 2000US-02330633P.
PR 14-SEP-2000; 2000US-02330644P.
PR 14-SEP-2000; 2000US-02330654P.
PR 21-SEP-2000; 2000US-02342233P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 25-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.

PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 08-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476159/51.
DR P-PSDB; ADM19833.
XX
PT Isolated nucleic acid molecule encoding a channel/transporter protein is
PT used in preventing, treating or ameliorating a medical condition.
XX
PS Claim 1; SEQ ID NO 161; 809pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule encoding a
CC channel/transporter protein or sequences at least 95% identical to a
CC these. The nucleic acids and proteins encoded by them are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. The antibodies to the proteins can also be used
CC in alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiodenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC The polypeptides can also be used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC also be used as a food additive or preservative to increase or decrease
CC storage capabilities. This sequence corresponds to a gene of the
CC invention.
XX
SQ Sequence 630 BP; 156 A; 163 C; 150 G; 156 T; 0 U; 5 Other;

Query Match 12.5%; Score 557.6; DB 5; Length 630;
Best Local Similarity 99.6%; Pred. No. 6.4e-158;
Matches 568; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
QY 3904 GCTAAAGGGGAATTTTGAGAGTCATTGAAATCTTATTGAAAGATGCCACAGATGTT 3963
DB 1 GCTAAAGGGGAATTTTGAGAGTCATTGAAATCTTATTGAAAGATGCCACAGATGTT 60
QY 3964 GTGGATCTCTCGTGAGGTTATGACATCATATGACTGCTTGAAGATCTTAGTT 4023
DB 61 GTGGATCTCTCGTGAGGTTATGACATCATATGACTGCTTGAAGATCTTAGTT 120
QY 4024 AAAAAGAAAGGCTTTCAAGAATGTTTCCAGGCATCTGCAGGTTCTACATGTCAGCTAT 4083
DB 121 AAAAAGAAAGGCTTTCAAGAATGTTTCCAGGCATCTGCAGGTTCTACATGTCAGCTAT 180
QY 4084 TATGAGCGGAATCACAGAATAGCAGTTGGAGCTCGCCATGGTTTCAGTGGCCCTGACGAC 4143

181 TATAGCGGATACAGATAGAGTTGGAGCTGCCATGGTTCACTGGCCCTGTACGAC 240
4144 ATCCGACTGGAAATGTGACAGCAATCCATGGGACAAAGGACCAATCACTGAGTGGCT 4203
241 ATCCGACTGGAAATGTGACAGCAATCCATGGGACAAAGGACCAATCACTGAGTGGCT 300
4204 TTTGCTCTGATGGAGATATCTTGCACCTACTCAACACACTGACAGCCACATTTCTTTT 4263
301 TTTGCTCTGATGGAGATATCTTGCACCTACTCAACACACTGACAGCCACATTTCTTTT 360
4264 TGCAGATGAACACGTCACCTGCTGGGAAGCATCGGCATGTCTGAATCTGGCACCCTCAGCTG 4323
361 TGCAGATGAACACGTCACCTGCTGGGAAGCATCGGCATGTCTGAATCTGGCACCCTCAGCTG 420
4324 CGCTGATTAACACCTACACAGTGGCCCTGTGACGCCCGCTGCCCGGCTCCCAAT 4383
421 CGCTGATTAACACCTACACAGTGGCCCTGTGACGCCCGCTGCCCGGCTCCCAAT 480
4384 GCCTCAAGTGGCCCGGCTCATCTGGACTTCCAAACCGCAAGTGCATCTCATGGGCCAT 4443
481 GCCTCAAGTGGCCCGGCTCATCTGGACTTCCAAACCG- AACGTCTCATCTCATGGGCCAT 539
4444 GACGGGAAGGAGCACCGCTTCATGGTCTAA 4473
540 GATGGGAGGAGCACCGCTTCATGGTCTAA 569

RESULT 10
ADM02366
ID ADM02366 standard; cDNA; 2578 BP.
XX
AC ADM02366;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human cDNA of the invention SEQ ID NO:1051.
XX
KW ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS Homo sapiens.
XX
PN EP1347046-A1.
XX
XX 24-SEP-2003.
PD
XX 12-APR-2002; 2002EP-00008400.
XX
XX 22-MAR-2002; 2002JP-00137785.
PR
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
PA
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-723558/69.
DR P-PSDB; ADM04809.
XX
XX New polynucleotides and polypeptides are useful in gene therapy, for
PT developing a diagnostic marker or medicines for regulating their
PT expression and activity, or as a target of gene therapy.
XX
XX Claim 1; SEQ ID NO 1051; 305pp; English.
XX
XX The invention relates to a novel human polynucleotide and the encoded
CC polypeptide. A polynucleotide of the invention may have a use in gene
CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC as a primer for synthesizing the polynucleotide or as a probe for
CC detecting the polynucleotide. The polynucleotides ADM0316-ADM03758 are
CC useful in gene therapy, for developing a diagnostic marker or medicines
CC for regulating their expression and activity, or as a target of gene
CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides

CC are useful as pharmaceutical agents. The present sequence represents a
CC cDNA sequence of the invention.
XX
SQ Sequence 2578 BP; 720 A; 531 C; 581 G; 746 T; 0 U; 0 Other;
Query Match 10.6%; Score 474.6; DB 11; Length 2578;
Best Local Similarity 55.1%; Pred. No. 3e-132;
Matches 999; Conservative 0; Mismatches 799; Indels 15; Gaps 3;
QY 29 TTGTTCTTTGGGTCGAAAGGCCCAACATGATCTCAGCGGTACTTTTAAACAGATG 88
Db 276 TGGCACTCTGGGACAGAGGCCCTCCCAACAGATCACTGCCATCATGATCACTGATG 335
QY 89 ATGGGGCCACAGATCGTAAACAGGATGTCACGCGGCAAAATATGTCTCTGGGATCTTTTCA 148
Db 336 ACACGGAAACGATTTGACTGGAAGTCAAGAGGTCAGCTCTGTCTCTGGAACTCTCAC 395
QY 149 TAGAACTGCAAAATTAATCTCGAGCACTGTGTTGTTGGTCATACAGCATCAATCACTGTT 208
Db 396 ATGAACATAAGATTTTCAGCGAAAGAACTCTATTTGGTTCATTCAGCTTCGGTAAACATGTT 455
QY 209 TGTCTAAAGCTTTGCTTCCAGTGACAAACAGATATATTGTGAGTGATCTGAAAGTGGAG 268
Db 456 TGGCAAGAGCAAGGAGCTTCTCTAAACAGCCCTACATTTGTTAGTGTCTGTAATAATGGGG 515
QY 269 AGATGTGCTCTGGGATGTGAGTGATGCGAGATGATTAAGAAATTTTCAAAATTTAGCTTGA 328
Db 516 AGATGTGCTTTGGAAATGTCACCAATGGACATGCGTGGGAAGGCTACACTCTCTTACA 575
QY 329 CACATACCTGGCATACAGTTCTTACCAGTTCTCTGTTGGGAATCAGCGAGAAGAGCTTT 388
Db 576 GGCACACTGCAATCTGTTATTACCACTGCTCGTTCGGGATCAGAGAGAGGCTGGCTTC 635
QY 389 TATGCCACGACATTTACCTGAAATCTTTGTTGGATGCTACAGGCTTTGAGTATTAAT 448
Db 636 TTTGTTGTGGAGATATCAAGATGCTCTTATAATTTGATGCCAAAACCTTTTGGCTGTGTTTC 695
QY 449 ACTCTCTAGTATCAAAGATATCACCAGACTGATTAAGTCTCCATGATGATTAATTTCCGATCC 508
Db 696 ACAGTTTATAGATCATCTCAGTTTCTGACTGGATCACTGCAATGTCATGTCATGTTCACTCCA 755
QY 509 ACCGAACACAGAGGACACAGTGGTAGCACTCTCGGTGATCGGCATCTCCGAAGGTGGA 568
Db 756 TGAGAAATTCAGAGAGATTTCTCTTTGGTGTATCAGTAGCTGCTGAGCTCAAAGTATGG 815
QY 569 TTTGTACCTCGGAAATAGTGAATGACAGGATCTGAGCAATATTTTGGAGGAGATCCA 628
Db 816 ATCTTTCTCATCTATCAACAGCAATTCAGGAAAGCAAGATGCTATGAAAGAAATCCA 875
QY 629 AACCAATTTATTTGTCAGAAATTCGCAAGCATCTCTTTTGTGATTTTACAAAGGTCA 688
Db 876 AGTTCTTGAAGTCTTGAATCTGCCAGCAATTCGATTTTGCATATATCTGAGAGACTTC 935
QY 689 TTTTGGTGTGTTCCTCAAAATATTGGAGGCTGTTCGATCCCGGAGACTATCTCTGTTGT 748
Db 936 TATTTGGTGTATTTTCTAAATGTTGGAAGGTTTATGATTTATGATTTTCCCTCTGTC 995
QY 749 GTTCAGGTCTTAGTGAATAATGGACAGACATGGAACCGGGGGGACTTTGTCTCATCAGATA 808
Db 996 TGACTGAAGTTAGTAGAAATGGGCAAGTCTTTTGTGTTGGAGAGTGTCTTAAAGAGACCACTTATC 1055
QY 809 AAGTCATCATTTGGACAGAAATGGCAAGTTATATTTTACAACTTACCTGCCAGTTGCC 868
Db 1056 GAATCTCATCTGGACAGAAATGGTCAAGTTATCATCTATCAGCTGCTGAACAGTGGGC 1115
QY 869 TTCCAGCTAGTGAATTCATTTCCGAGTGTGGGGAAGGAGTGAATAATTTAATTTCTTC 928
Db 1116 TTTCAAAAG-----CATATACCTGCTGATGGAAGAGTGTCTTAAAGAGACCACTTATC 1169
QY 929 CTGTACACATATCTCTCTTGGATCGAAAGATTAAGAGTTGCTTAATTTTCTCTCTCTTTA 988
Db 1170 CTCATTTACTGTGCTCTACTTCTGTGTCAGGAAATTAAGAACAGAGCCGCTCTCTTTGTTA 1229

```
QY 989 CTCGGTCTTCTATGATGACAGAGATATTTCCATAAACTGTTAAATTCAGGGTGATCTT 1048
Db 1230 TGGCTACATGAATGAAGGAAGAGAGCTTTTATCAAGGTACTTTTCTCGGAGAGTCT 1289
QY 1049 CTGGAAGCTTGAATATTTGGAACATATCAGACACAGCTG-----ATAAACAGGAGAGT 1102
Db 1290 CAGAGAGATTAATTTGTCGACATCTCTGATGTTCTGTATCCAAAGTTGATGGTCTC 1349
QY 1103 AAGAAAGCTGGCAATCAACACTCTATTAGTTTGCAGAGGCAATTTGATAAACTGAATC 1162
Db 1350 CTAGAGAGATACAGTAAGTACCCACCTGGACTCTTCAAGATAATTTTGTAAAGCATGATA 1409
QY 1163 CTTGTCTCTGCGAATTTATAGATCAGCTGAGTGTGATTCCTCAATAGTAAGAACCTCTTA 1222
Db 1410 CTATGTCAAAAGTATTATTGACTATTCTCTGGGCTTAAAGATGGGGCAGGAAGTCTG 1469
QY 1223 AAGTAACCTGCAAGTGTGTACATACAGACATGACGACTTGTGTGGTCTGTGAAGATG 1282
Db 1470 TAGTCACTTATCAGAGTATATTCAGTCTTGTAACTATATGTTGGCTGTGAAGATG 1529
QY 1283 GAAGCATAGTTATTGTACCTGCCACACAGACGGCCATAGTACAGCTGTTGCCAAGGGGAA 1342
Db 1530 GGACAAATTATCATPACCAGGCTTTGAATGCTGCCAAAGCAAGACTTCTGGAAGTGGTT 1589
QY 1343 ACATGCTCAGAGAGGTTGGCCACTCAGACAGACACTCCGTGTATCGGAAACAAGTCA 1402
Db 1590 CTTTAGT---AAAAGATTCTCCCTCATAAAGTTCTTAAAGGCCACACCAAGTGTCA 1646
QY 1403 CATGTTTGTATATCTCATCAGCTCTCAGCTGGTATGATCAAGATACCTGATATCTG 1462
Db 1647 CTTCAATTACTTACCAATGCTCTCTTGGAAATTAGACCAAGTTGATGTTGTCTG 1706
QY 1463 GAGGTGTGGATTTTTCAGTCATATTTGGACATATTTCTGGAGAAATGAAACATATCT 1522
Db 1707 GGGACCTGGACTCATGTGATCTTGTGGGATATCTTTACTGAAGAAATTTTGCATAAT 1766
QY 1523 TCTGTGTTGATGTTGGTGAATTAATCACTTCTAGTTTCCACCTGAAACCTGATGTCAA 1582
Db 1767 TCTTTTGGAGCTGGTCCAGTAACAAAGTCTTTTGTATGTCACCAAGAGATTTAAACTAA 1826
QY 1583 GAGTACAGCACTGATCTCTGTAGCCAGTACCACTCAGTAGGACTTCTAAGTTTGC 1642
Db 1827 GGGGTGAGCAGATAATTTGTGTGTGTGGGTGACCAATCCGTGGTCTCTTCCACCTTG 1886
QY 1643 GAGAGAAAAATGATTAATGTTGGCATCTCTCACCCTTTTCTTATTTCAAGTAATCAAA 1702
Db 1887 AGGAAAGAGTTGCTCTGCATGCCCGGAAGACACCTTTTCTGTGAGGATGATAAAT 1946
QY 1703 GAGGCGCTTCTGATGATTAATCTGTGTGGGTGTTGAGATGTTCTGTGTTGTTGTTG 1762
Db 1947 GGCACCCGGTTGAGAAATTTTAAATTTGATGTGAGATGATCACTCAGTTTATATCTGG 2006
QY 1763 AAATGGATCTGGTGCATGCTGTTGTGTGATGGGATAACAGCAGTTGAGATTTCTAA 1822
Db 2007 AAATTGAAACAGGCACCTTTGGAAGAATGAGACAGGAGAGAGACGCAATATTCTTA 2066
QY 1823 ACGCTTGTGATGA 1835
Db 2067 ATTCTTGTGATGA 2079
```

RESULT 11

ADS96553

ID ADS96553 standard; cDNA; 4578 BP.

XX

AC ADS96553;

XX

DT 02-DEC-2004 (first entry)

XX

DE Drosophila melanogaster protein coding sequence, SEQ ID 174.

XX

KW Insecticide; Antiparasitic; Anthelmintic; gene; ds.

XX

OS Drosophila melanogaster.

XX

PN WO2004039999-A2.

XX

PD 13-MAY-2004.

XX

PF 08-AUG-2003; 2003WO-US024982.

XX

PR 30-OCT-2002; 2002US-0422377P.

XX

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX

PI Stam L, Kandar KP, Spana E, Bachmann J;

XX

DR WPI; 2004-376203/35.

XX

DR P-PSDB; ADS96554.

XX

Identifying a compound that inhibits the activity of a protein for Drosophila viability for use e.g., as insecticidal agent by expressing in a recombinant host a DNA molecule to produce a protein essential for Drosophila viability.

XX

PS Claim 1; SEQ ID NO 174; 57pp; English.

XX

The present invention relates to a method for identifying a compound that inhibits the activity of a protein essential for Drosophila viability. The method comprises: (a) expressing in a recombinant host a DNA sequence encoding a protein essential for Drosophila viability; (b) testing compounds suspected of having the ability to inhibit the activity of the protein expressed in (a); and identifying a compound tested in (b) that inhibits the activity of the protein. The method is useful in identifying a compound that inhibits the activity of a protein essential for Drosophila viability for use as insecticidal, ectoparasiticide, antiparasitic, anthelmintic or acaricidal agent. The present sequence is the DNA sequence for one such protein essential for Drosophila viability. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 4578 BP; 1053 A; 1281 C; 1303 G; 941 T; 0 U; 0 Other;

Query Match

Best Local Similarity 8.8%; Score 393; DB 13; Length 4578;

Matches 2153; Conservative 0; Mismatches 2275; Indels 195; Gaps 20;

QY 1 ATGGCAGGAAACAGCCTGTTCTTACCCATTTGTTGGGTGCGAAAGCGCCACACAT 60

Db 1 ATGGTGAGTACCATCTGGTGTACCCGTGGTGTCTGTGGGTCCACCGCGCCACGCAC 60

QY 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAAACAGGATGTCAGC 120

Db 61 TGCATCTCAGCGGTGTTCTCTCGGACGATCAGTTTACCTGTGTACCGGCTGCTAGGAT 120

QY 121 GGACAAATATGTTCTCGGGATCT---TTCAAGTAGAATCTGCAATTAATTCCTCAGCAGCT 177

Db 121 GGACAGATCTGCTTTGGCAAGTGGAGCCACACCCCTGGAAGATGTGCGCGGTGCTC 180

QY 178 TTTGTTGGTTCATACAGCATCAATCACTGTTGTTGTCTAAAGCTTGTCTTCCAGTGACAAA 237

Db 181 TTGGTGGCCATTCGGCGCCCGTTTGTGCTTAGTCGCTGCTCGCTTCTGCGGAGAAC 240

QY 238 CAGTATATTTGTGAGTGCATCTGAAAGTGGAGAGATGTGCTCTGGGATGTGATGTGCG 297

Db 241 AACTTCTCTGTGAGTGTCTGAGAACCGAGAGATGTGACCTGGGATCTTACGGACGGC 300

QY 298 AGATGATTAATTAACAAAATTAGCTTGCACACATCTGGGATGATAGTTCTACCAAGTTC 357

Db 301 AAGTGCATGGAGCGGTCAAGCTTCCACAGGTGCACACCAAAATTCAGAGTACCAACAC 360

QY 358 TCTGTTGGGAATCAGCGAAGAGAGGCTTTTATGCCACGAGCATTTACCTGAAATCCTT 417

Db 361 -----GCCAACAGCAGGAGATGTGGCCCTCTTTTGCATTTGGCTACTATCTGAAATCATG 414

Qy	418	GTTCGTGATGCTAC	CACGCGCTTGAAGTATATATAC	TCTCTAGTATCAAAAGATATAC	CCAGAC	477	
Db	415	GTCA	TGATCCGTTT	CAGCTAGAGGTGATCTAC	GTGCTCAGCTCAAAAGGTCAAAACCGGAT	474	
Qy	478	TGGATTAGCTCCAT	GAGTATTTATTCGATCCCA	CCGACACA	CAAGAGGACACAGTGTGTAGCA	537	
Db	475	TGGATATCCGCTAT	TTCACGTACTCGCGCCCAT	TGCGTCAAGGATGAC	GTAGTGTGTGGCC	534	
Qy	538	CTCTCGGTGACT	TGCGCATCTCTG	AGGCTCTGGATCTTAC	CTCGGAAATPAAGTGATCATGCG	597	
Db	535	ATCAC	CCACCGGAA	CCGTGAAGCTCTG	AGCTCTGAC-----GGGTACAGAGACAA	588	
Qy	598	GATACTGAGCCAA	TATTTGAGGAGGAAT	CCAAACCAATTTAT	TGTTCAGAAATGCCAAAGC	657	
Db	589	CACGCGAGCCCA	TTTACGAGAACGAAT	CCAAGGAGATCCG	CTGCTTGAATGCCATCACC	648	
Qy	658	ATCTCTTTTGTG	CATTTACAAAGGTGACTTTT	TGTGTGTGTTCCAAATAT	TGTGGAGG	717	
Db	649	ATGAATTTGCT	CGCGCCAGAAC	CAGCGACAGTGTCTTTAGTTTGTAC	CAAAATATGTGCAG	708	
Qy	718	GTGTTTCGATGCG	CGGAGACTATTCCTTGTGTGTTC	TTCAGGTCTCTAGTGA	AAATGACACAGA	777	
Db	709	ATCTACGATGCG	CGGACTTCAC	TGTACTCTGCTCCGTAAT	TGCTCCGCAAGGGAACGT	768	
Qy	778	TGSA	CCGGGGGGGACTTTGTCTCAT	CAGATAAAGTCA	TATTTGGACAGAAAAATGGCAA	837	
Db	769	TGGCAGGCGCG	AGCTTCATCAC	CTCCGACCGGTAATGCTAT	TGGACAGATGAGGGAAG	828	
Qy	838	AGTTATATTTT	ACAACTAC	TGCGCAGTTTGCCTCCAGCTAGT	TGATTCATTTCCGCACTGAT	897	
Db	829	GGCTACTTGT	TACAAGCTGCCCGCAAACTG	CAATCCGACAA	CAGGAGTTTCCA	888	
Qy	898	GTGCGGAAGGC	AGTGAATAATTT-----TAATTCCTCTGT	TACAAATATCCTCTTGGATC	952		
Db	889	AGTGTGGT	CAGGATGACCTTTACTCTA	CTACGTGTGTGCA	ACGCGGTGATPAAGGTG	948	
Qy	953	GA	AAAGATAAAGAGTTGCTAATTTGT	TCTCCTCTGTACTCGGTTCTCTAT	TGGATGCGAG	1012	
Db	949	CTGTCTG	CCCCCAGCCATGAAGTTGCTT	CAGGCGCGGTG	SCCAGCAATCTACTG	1008	
Qy	1013	AA	TATTTCCATAA	ACTGTTAAATCAGGGTGATCTTCT	TGMAAGTTGAAATATTTGGAA	1072	
Db	1009	CGCGAGACT	CCGAGGATACATCTCTGT	TGGAATGTGCGGAGGTG	CCGTTAGACAAT	1068	
Qy	1073	TATCAGA	CACAGCTGATAA	CAGGGAAGTGAAGAAGGCTGG	CAATGACAACT-TCTATT	1131	
Db	1069	ATTAGCATAT	TCGAGGCCAAACAGATGCG	CGCGGACCCCTCA	AGCGCAGCTGTGCACT	1128	
Qy	1132	AGTTT	TGCAAGGCAATTTGATAAACTG	AAATCTTGTCTGTCTG	TGGAATATATAGATCAGCTG	1191	
Db	1129	TCGTTTGGT	TCGAGGCGTGGTCCATAT	TGATCCACCA	CCAGTGGSCATCCTGGATCAGC--	1186	
Qy	1192	AGTGTGAT	TCCCAATAGTAT	ATGAACCTTTAAAGTAA	CTGCAAGTGTGTATACACAGCA	1251	
Db	1187	----	TCTCCCGCAT	CACAGAATCGCCCGGTGAAGCT	CACCTCGAGCATCTATCTGCCACAG	1242	
Qy	1252	CATGACG	ACTGTTGTTGTCGTG	AGATGGAAGCATAGT	TATTTGTACCTGCCACACAG	1311	
Db	1243	CAGAGT	CGCCTGGTATCG	AGCGGAGATGGCAGCATAGT	CTATGTCAGCCACCCAG	1302	
Qy	1312	ACG	GCATAGTACAGCTGTTT	GCAAGGGGAA	CA	TGCTCAGAAAGAGGTGTGGCCACCTCAC	1371
Db	1303	ACAGT	GATGATGCAGCTGCT	TGTTGGCATCAAG	CAAAACTTTCACCGATTTGCCCTCGCAT	1362	
Qy	1372	AGAA	CATCCGTGGTGCAT	TCGGAACAAAGT	CTACATGTTTGTCTATATCTCTCATCAGGTCTCA	1431	
Db	1363	CAGAT	CTCTCTACGGTGCAT	CTGTGGAAGGGTCAA	TTTGCCTGTCTCTGCCCTCGATGATTTAC	1422	
Qy	1432	GCT	CGTATGATCAAG	GATACCTGATATCTG	AGGTGTGGAATTTTTTCAGTCAATAATTTGG	1491	
Db	1423	TCG	AGATATGAGA	AGTGCATCTACTTTTC	CGCGGCGCATCGAATTTTCGCGGTGTGTCTGTGG	1482	
Qy	1492	GACATATTTT	CTCGGAGAAATGAA	ACATATCTTCTGTGTT	CATGTGTTGTTGATGATTTACTCAA	1551	

Db 2563 AGACCCATTGTGCCATCTCATACGGAGTGTCTTCAAGCCGGCTATATGTGCTGCTG 2622
Qy |||||
Db 2542 CTGCGGGTTATAA-----TCAGCTGCTTGTAAACTGTAC 2578
Db 2623 CTGCCACATGGCAGAACTATGCCATACCAACCCGGTATCCAGCTTCCTCCAGTTCC 2682
Qy 2579 ATGGGAAAACAGAGTAGGAAG-----GAACTGCGAGGCTCTGAGGAGTA 2625
Db 2683 AAGAAGCGCCGCTGCCAGAGAGCTGCAGCGCTTGGAGCATCTCAGCGCTGTTTTACA 2742
Qy 2626 GGAAGGGAATTTACGGAGTGTCCGCTGCGGTCAACACACAGCATCTCTGTCTATCAT 2685
Db 2743 TCCGCTCTGATTTGGAGCTCAGCACTACGTTGACCAACCAATATCTTGGCTGGTG 2802
Qy 2686 TCTTTGGCAATATCTTTAATAGATATGACCAATGCAATTTTATTTGGTATCATATGAAG 2745
Db 2803 GCCATGTCCAATACGCTGTCTGATGAGTGGGCTCTCTTTCTGCGGACACGAGAG 2862
Qy 2746 AAGGTCCTACAGGCCACCTAGACCAACAGCACCCAGACCTTTCTAAGGCAAGGGTTCC 2805
Db 2863 CACAAGAGCTGCAG-CGCTTGGCCAGCGCACGGAATCCACATTTGAGCAACGAGGAGGA 2921
Qy 2806 CCTCCAACTTCCAGTAATATTTGTGCAAGGACAGATTAAACAAGTTGCTGCACCTGTCTT 2865
Db 2922 GCGGAGGAGCTGATGCCCCACCACTCTCTCAATCAAGCAGCCTGGAGCCTGCTTGC 2981
Qy 2866 TCCGCTGGTCTGATGCTGATCACTCTGGCTCTGACCTCTCTTCTGCTCTGCTTTACAT 2925
Db 2982 TACCCACCATTTGCTTCTGTTGCCGACAAATTTGAGGCCCTGGAGCCGAGAAATTCAA 3041
Qy 2926 ACTGTTTCTTAGTAATGAAGTTGGATGATGATGCTGCTGATGCTGATGCTGTTTGTCTG 2985
Db 3042 GCCTCCGCGAGTGGAAATGATGTCAAGCGCTGGCAGCATCACTGCTATAGAG--ATCAGG 3099
Qy 2986 CCAGACTACTGGATTTGGATAATTTAGGCTCTCCCTCTCTGAGATGCTGGCCCAAGA 3045
Db 3100 GAGCGGCTCAGCAGATCTCTTTTGGGGAGCTTACCCGATGGGCAAGAGGTTGCCAAG 3159
Qy 3046 TGGCAAGATCGATGCTTGGAGGTGAGAGAGCGGCACAGGCCCTGCTTCTGGCGGAATG 3105
Db 3160 CAGCTAGTCGAGAGCTGGGCACAAATATCTGCCACTCTACACGACACCGAGCCCATAGT 3219
Qy 3106 AGAAGATTTGAGCAGGACGAGGAGGAGGAGCCATGATGCTGGGCTCTTACTTACCT 3165
Db 3220 GGAGCCCAACAGCAGTTGGCTTTGATAAGTTCAGCGCGCAGCGAGGAGTGGCTCGGA 3279
Qy 3166 CAGTACATACACCACTATCATCACCTGGAGTCACATCAGAGCCGCGAGACTATCAC 3225
Db 3280 TCGGCGGCAATGGAGGGTGGGCTCGGAGTATCCGAGGCGGTGGTGTGTTAGTGT 3339
Qy 3226 ACGGCTCTGATGCTCAGGCTTGAAGCAAAAGTCCAGGAGGAGAGCATGACCTTGT 3285
Db 3340 AGCGGCGCAGTGGCAGCTTCTTGGGGAGATGCCCAACCAAGATGAGACTACGAGAG 3399
Qy 3286 GAGGATGACATACCACTGTTGCTTATCAAGTGTCCCAAAATGAAAAATTTCTAC- 3344
Db 3400 GAAGAGGAGGAGATTATACGAAACCATCCAGCTTGTGAGCTGTAAGCGCAAAACAAT 3459
Qy 3345 -----ATCTTACGAGGAAGACGAGAGCAGCTACCGCTATTGTTTACTTGGAGT 3395
Db 3460 ACGGAGTTATCTTACTGCTGTTAATAGGAGCAGAAATTTGGCCAGGATATATCCAGGAG 3519
Qy 3396 AATAGGAGCTGAAATTTG--GTGCTGAAATTTGAACCTCTTAAACTATTGACCAAGACCTCGA 3453
Db 3520 TCACCGAATCACCGTGGAGTATTAGCATGGCCACGAGGAGCAATCTTACAGTGGAGTT 3579
Qy 3454 AGCTTAGCCAAATCTCTAGGAGATTGGGTTGACTAGTGGTGGATGCCAACTACTCGCTG 3513
Db 3580 GCTGGGGGAGAGCTCGAAATCTAGTGTCTGGAGGGCTTCGGCATAGCAACAATCTA 3639
Qy 3514 GCCAGACATCTTGAAGGACCTGAGCTTCTTCTGCTACAGCCTCAAGCCCAACTT 3573
Db 3640 GCCCGTCTTACTTCAATTTGGCCCTGGCTCACTTGTCTGACGCGCACCATCTCCCAAGCTG 3699

Qy 3574 CTTCCACACAGCAGCTATCCGAAGAACAGCAATGTGATCTGATTGGACGCTGGTTCACTGTT 3633
Db |||||
Qy 3700 CCGCAATACACGCCACTGCGCGGCGAGCATCGATCTGTGGGTCTGGATTCACCGTT 3759
Qy 3634 TGGGAGCCTTACATGATGTGTCCGCTGTTCTGATGGGCTTCTCGAACTTTGTGCGGAT 3693
Db 3760 TGGGAGCCGTACTGATGTGTCCAAAGGTGCTCTTTGTTGCTGGAGATATCGTCCGAG 3819
Qy 3694 GCCGAGAAACAATCTGCCAAACATCAATATGGGTTGCCCTCTGAGCCCGACAGCTGACTCG 3753
Db 3820 GGC---AAGAGCATGCCAATCTGAATCAAACTTCACTAACCCCGCAGCGGATGCC 3876
Qy 3754 GCGCGCTCTGCGAGCATGCCCTCTGCTCATTTGCCACCGCCAGACCAACCCGCTTTCATC 3813
Db 3877 TGCAGAACCCGCGAGCATGCTCTGCGCTTGTATAGCCACGCTCGCGCGCGGCTTTCATC 3936
Qy 3814 ACCACATAGCCAAAGAGGTACACAGACATACGGCTCTTGCAGCAATATCCCAATCACAG 3873
Db 3937 ACTGATGGCGCGAGGTAGCCAGGTACAATACGATGCAGCAGAACCGCAGTCCATC 3996
Qy 3874 CAGAATATGCACACAACAATCTTTGCAC---GAGCTAAAGGGGAAATTTTGAGAGTCAAT 3930
Db 3997 AACACGCCGCTCACCCAGTCCGTTCTGCACAGGCCACAGGCGAGATTTCTCAGTCCGTT 4056
Qy 3931 GAAATTTCTTATTAAAGATGCCACAGATGTTGTGGATCTTCTCGTGGAGGTTATGGAC 3990
Db 4057 GAGATGCTAATCGACAAATCGATCGGAGATCGCGGCTCTCTGTGGAAGTTATGGAC 4116
Qy 3991 ATCATATGATGCTGCTTGAAGATCTTTAGTTTAAAGAAAGGCTTCTCAGAAATGTTTC 4050
Db 4117 ATTGCCCTGACCTGCGTGAGCGGCAACGAGCTGAAAGAACCGCGGCTTGGCGAGCTGTGT 4176
Qy 4051 CCAGCCATCTGCAAGTCTTACATGCTCAGCTATTATGAGCGGAATCACAAGATAGCAGTT 4110
Db 4177 CTGCGCATCTGCAAGTTTAAACAGATCTCGCATCTGCGCCCAACACGCGCATCGCCGTG 4236
Qy 4111 GGAGCTCGCCATGTTTCAAGTGGCCCTGTACGATCCGACATCCGACCTGGAATAATGTCAACAATC 4170
Db 4237 GCGGCCAACAGTGGCAACCTGGCCATTTACGAGCTGCGCCAGAACCAAGTGCCAGATGATC 4296
Qy 4171 CATGACACAGGAGCAATCACTGAGTGGCTTTTGTCTCTGATGGAGATATCTTGC 4230
Db 4297 CCGGCCCAACAGCATCCGATCACTCTGCTGGCTTTCTCGCCGATGGCAAGTACCTGGTC 4356
Qy 4231 ACTACTCAAAACACTGACAGCCACATTTCTTTTGGCAGATGAAACAGCTCACTGTGCGA 4290
Db 4357 TCGTACTCTTGGCGAGAGAACCGCTCTCTTCTGGCAGACCTCAACGGGATGTTCCGA 4416
Qy 4291 AGCATCGGCATGCTGAATCTCGGCACCTCAGCTCGCTGATTTAAACCTTACCAGGTGCC 4350
Db 4417 CTGGGC-----CAGTCCGACAGCGCTGTACGAAGGCTACTCTACGGCA 4461
Qy 4351 CTGTGTCAGCCCGCTTCCCGCGCTCCCAATATGCCCTCAAGCTGGCCGCGCTCATCTG 4410
Db 4462 CCCATTCGGGATGTGTGCG-----GCCTGAATCCGATCGGCTTGGCCAGCTTGTGTG 4515
Qy 4411 ACTTCCAAACCGCAAGCTCATCTGATGCCCATGAGCGGAAGAGCAGCCCTTTCATGTC 4470
Db 4516 ATCAAAATTCGACCGGTGACCTCATGCTGGCCGAGGCTCCGAGAGCGGATTCATATGTC 4575
Qy 4471 TAA 4473
Db 4576 TAA 4578

RESULT 12

ABL19261

ID ABL19261 standard; DNA; 5221 BP.

XX ABL19261;

AC ABL19261;

XX DT 26-MAR-2002 (first entry)

1492 GACATATTTCTGAGAAATGAAACATATCTTCTGTGTTCTATGTTGAGATTAATCA 1551
2126 GATCTGTACAGCGAAGCTTACTGCATAGGTTTTCGCTTTCATGCGGTGAGATCACAAA 2185
1552 CTTCTAGTTCACCTGAAAACCTGTAGTGAAGATACAGCACTGCATCTGTGTAGCC 1611
2186 CTGCTGTGCGCCAGAGAGCTGTAGTCCAAAGATCTGAAATGCAATATGTTCCGTGCC 2245
1612 AGTGACCACTCAGTAGAGACTTCTAAGTTTGGAGAGAAAAATGCAATAATGTTGGCATCT 1671
2246 TCGGACCAATTCCTGTTACGTTAGTGTCCCTTCAGGAGCGCAAGTGTGACCCCTGGCAAGC 2305
1672 CGTCACCTTTTCTTATTCAGTAATCAATGAGGCTTCTGATGATTACCTGGTGTG 1731
2306 AGACATCTATTCCTGTGGTTACATTAATGAGGCTCCCTGGACGACTTCTTAATCGTA 2365
1732 GGGTGTTCAGATGTTCTGTGTACGTCTGGCAATGATATCTGTGCATTCGTTGT 1791
2366 GGCTGTTCCGATGGCAGGTGTATGTGTGGCAATGGAACGGGTCACTCGATCGAGTG 2425
1792 GTGATGGGATAACAGCAGTTGAGATTTCTAAACGCTTGTGATGAA 1836
2426 CTTTACCGGATGCTGGCAGAGGAAGTCTTTCGCTTTCGATGAAACAGGCGGAGGATGC 2485
1837 -GCTGTTCTGCTGCTGTTGATTTCACTTAGTCACTCAGCAGTCAACCTAAACAGCTAT 1895
2486 GGCTCGGGCGCGCGGTGTTCAATGGTGCATCTGCCAGGAGATGGGTATGGCCAAAT 2545
1896 GACGAGACGTAGTCTTCCTGCTCTTAAATAATGCG 1941
2546 CCGGAGTGCATCTTCTTCGCTGGCTGAGTCCGCGAAATGATGCAATCCGTCATGCT 2605
1942 CTACAAACCTTGCACATCACTCTTGGCTTCTGAGGCATCTGACAGGGAAA 1994
2606 ACCCAGCGCGGATTTACGCAACTCCAGCACTTCCAGGTCATAACCAAGGAACTTCGAC 2665
1995 --TTTACCTAAATTTCTCATTAATCTCCTGATGTTCAAGCAATAAAGCAAACTTAACA 2052
2666 TTTTCTGATGAACACCGCAGCAATCTTTAGTGTATCCAGGCTCTTCGTACTAATCTCAAG 2725
2053 GACCGGACATACATGTGCTATTCTTTGATGTGAGCGTTGATTTCAACTCTCTGACT 2112
2726 GACGGGAGGACACATACCTTTCTTGAATCGAGGACTGATCTTGAATTTGCATAGC 2785
2113 GAAGAAGCCTTAGGCGGAAATACGTCTTTATTTCCAGAGAAATTTGCAAAAGCATCT 2172
2786 GAGGAATACGACAGATGACACCACTACCTTGGATCACTTGTGTCCACCTGCAGAT 2845
2173 GGCAGTTCAGAAAGGGGCTCTTTTAACTGGAAACGAGCAGAGTCTCTTCCAA 2232
2846 CCAGAGATGGCAATTCGATGCAATTTGGATGCTAGCAAGAGATCGGTGACTTCTTCAAC 2305
2233 CAAGTGAAGAAACGATCAAGAGAAACATCAAGAAACACCTCTTGTATGATGAAGAGGAG 2292
2906 AAGTGAAGAACAGCGGTAGATGTGGAGAGATCTTGAAGCAAGAGCAACAAACCGGG 2965
2293 GATGAGGAGATAATGAGGCGAGAGAGGGA 2321
2966 CTGGTCAGAAATTCAGGAGAAACTGAGATTTGCGAGAGAAAGTACAGGCCAAGGTG 3025
2322 -----AGAAAGTGATCTGAAATATCGGTCCAGCAAAATCAAAAGCCA 2361
3026 GAGAGCCTGCAAAAGGCGAGTAGAGCGCGCAGGAGCAGCAGGACCTGAAGAGCAAAATC 3085
2362 TTGACCTATTAGAAATATTAATTAATGATGCACTGCAAGCTGTTATGTTCTGCTT 2421
3086 GCTTCCAAAGATGAGGTTACGATGTATGAGGTAGCCAGCTACTTCTTTCGCTGCTC 3145
2422 CAGCCTTGGGGTTGAATGAAGTACTGGATGAAGTTTGCCTGATCGCCTTGGAAATGCTG 2481
3146 CATTTCTGGGGCTAGATCCACATCTAGACAAAGATGTGAGAGCGGCTTGGGCTGCTT 3205
2482 AAACCCACATGCAACCGTATCGTTTGGCTCTTGTCAAGAGGAGGCCATATGTCATGATG 2541

3206 AGACCAATTTGTGCCCATCTCATCGAGTGTCTTCAAGGCCGCTATATGTGCGTGTG 3265
2542 CTGCGGGTTATAA-----TACGCTGCTTGTAAACTGTGCAC 2578
3266 CTGCCCACATGGCAGAACCAACTATGCATACCAACCCTGGTATCCAGCTTCCCTCCAGTTCC 3325
2579 ATGGGAAAACAGAAAGTAGGAAG-----GAAGTGCAGAGCTGTAGGGAGTA 2625
3326 AAGAAGCGCGCTGCGCAGAGAGCTGCAGCGCTTGGAGCATCTCAAGCTGTGTTTCA 3385
2626 GGAAGGGAACCTTACGAGTGTCCCGTCCGTCACCAACAGCATCTCCTGTCTATCATTT 2685
3386 TCCCGTCTGCAITGGGAGCTCAGCACTAGCTTACCAACCAACCATATTTCTTGGCTGTG 3445
2686 TCTTTGGCAAAATCTTAAATGATGATGACCAATGCAACTTTTATTTGGTGATCATATGAAG 2745
3446 GCCATGTCNAATAGCTGTGTCGATGAGTGGCGCTCTCTTCTGCGGACAGGAGAG 3505
2746 AAGGGTCTTACCAAGCCACTAGACCAAGCAACCCAGACTTTTCTAAGCAAGGGTTC 2805
3506 CACAAGAGCTGCAG-CGCTTGGCCAGCGCAGGACTCCACATTGAGCAACGAGGAGGA 3564
2806 CTTCCAACTTCCAGTAATATTTGTGCAAGGACAGATTAACAAAGTTGCTGCACCTGTCTT 2865
3565 GCGGAGGAGCTGATGGCCCAACCATCTCTCAAAATCAAGCACGCTGGAGCCTGTCTTC 3624
2866 TCCGCTCGTCTGATGCTGATCACTCTGCTCTGACCTCTCTTCTGCTCTCTGCTTACAT 2925
3625 TACCACCAATGCTTCTGTTGCCCGCAAAATTTGAGGCCCTGGAGCCGAGAAATTCAA 3684
2926 ACCTGTTTCTTAAATGAAGTTGGAGTTCAGTTAGTCTATGCACTGTGTTATGCTG 2985
3685 GCGTCCGAGGTGGAATGATGTCAGCGCTGGCAGCATCACTGCTATAG--ATCAGG 3742
2986 CCGACCTACTGGGATTTGATTAATTTAGGCTCTCCCTTCTTGGAGATGCTGGCCGAGGA 3845
3743 GAGGCGGCTCAGCAGATCTCTTTGGCGGAGCTTACCCGATGGGCAAGAGGCTGCGAAG 3802
3046 TGGCAAGATCGATGCTTGGAGGTGAGAAAGCGCACAGCCCTGCTTCTGCGGGAACCTG 3105
3803 CAGTACTGAGAGCTGGGCAACATATCTGCCACTCTACGCAACACCGAGCCCATAGTG 3862
3106 AGAAGAAATTTGACGAGCAGGAGGAAAGCAATGATGCTGCTGCTCTTACTTACCT 3165
3863 GAGGCCCAACAGCAGTTGGCTTGTAGTCAAGCTCAGCGCGCAAGCGGAGAGCTGCTCGGA 3922
3166 CAGTACATAGACCACTCATATCACTGAGTCAATCAGAGCCCGCAGACTATCAAC 3225
3923 TCGGGCGCAATGGAGGGGTGGCGTCCGAGTATCCGAGGCGGTGCTGCTGCTAGTGT 3982
3226 ACGGCTCTGATGCTCAGGCTGAAAGCAAAAGTCCAGAGGAGAGCATGACCTTGT 3285
3983 AGCGGCGAGGTGGCAGCGTTCTTGGGGAGATGCCCAAGATGAGATGAGACTACGAGAG 4042
3286 GACGATGACATCACTGCTTGTCTTATCAAGTGTCCCAAAATGAAAAATTTCTAC- 3344
4043 GAAGAGGAGGAGATTATACGAAACCATCCAGCTTGTGCGAGCTGAAGCGCAAACT 4102
3345 -----ATCTTACGAGGAAAGCAGGAGAGCTACCGTATTTGTTTACTTGGAGT 3395
4103 ACGGAGGTTATCTTACTGCTGTAAATAGGAGAGAAATTTGGCGCAGGATATATCCAGGAG 4162
3396 AATAGAGCTGAATTTG--GTGCTGAATTTGAACTCTTAAATTTGACCGAGCTCGA 3453
4163 TCACCAATCAGCGTGAAGTATTTAGCATGGCCACGGGAGCAAAATCTAACCTAGTGAGTT 4222
3454 AGCTCTAGCAAAATTTCTCAGGGATTTCCGTTTGACTAGTGTGGATGCCAACTACTCGCTG 3513
4223 GCTGGCGGGAACGTCGAAATCTAGTGTGCTGGAGGCTTCCGCAATAGCAACAACTCTA 4282
3514 GCGACATACCTTGCAGGCACTGAGCTTCTTCTGCTACAGCTTCCAGGCCCAAACTTT 3573

Db	4283	GC	CGCTCTTACTTCAATCGGGCCCTGGGCTCACTTGGTGTGACGGCCACCAATCTCCCAAGCTG	4344
Qy	3574	CT	CCACACAGCACTATCCGAAGAAACAGCCATTGATCTGATTTGGAGTGGGGTTCACTGTT	3633
Db	4343	CG	CAATACACGCCACTCGCCGGCAGCCATCGATCTGTTGGGTGTTGATTTACCGTT	4402
Qy	3634	TGG	AGCCTTACATGAGATGTTCGGCTGTTCTGATGCGGCTTCTCGNACTTTTGTGCGCAT	3693
Db	4403	TGG	AGCCGTAACCTGAGATGTGTCCAAGGTGCTTCTTGGTTTGTCTGGAGATATCGTCCGAG	4462
Qy	3694	GCC	GAGAAACAACTTGCACATCACAAATGGGGTTGCTCTTGAGCCCGACAGCTGAATCG	3753
Db	4463	GGC	----AAAGCATGCCAAATCTGAATCAAACTTCACTAACCCCGCAAGCGGATGCC	4519
Qy	3754	GCC	CGCTCTGCGAGGCATGCCCTCTCGCTCAATTTGCCACCGCCAGACACACCGGCTTCAATC	3813
Db	4520	TGCC	GAACGCCAGGCATGCTCTGCGCTTGATAGCCACCGCTCGCGCGGGCGGTTCAATC	4579
Qy	3814	ACCA	CCATAGCCAAAGAGGTACACAGACATACGGCTCTTCGAGCAATACCCAAATCACAG	3873
Db	4580	ACT	ACGATGGCGCGCAGGTAGCCAGGTACAATACGATGACGACGACGCGCAGTCCATC	4639
Qy	3874	CAGA	ATATGCACACAACTCTTTGACAC--GAGCTAAAGGGGAAATTTTGGAGAGTCATT	3930
Db	4640	AAC	AGCGCTCACCCAGTCTGGTCTCTGCAAGGCCAAGCGCGGAGATTTCTGCAATGCGTT	4699
Qy	3931	GAAT	TTCTTATGAAAGATGCCCAAGATGTTTGTGGATCTTCTGPGAGGTTATGGAC	3990
Db	4700	GAG	ATGCTAATCGACAAATGCAATGCGAGATCGCGGCTCTCTGTGTGAAGTTATGGAC	4759
Qy	3991	ATCA	TATGTACTGCTTGAAGGATCTTTAGTTTAAAGAAAGGTCTTCAAGATGTTTC	4050
Db	4760	ATT	GCCCTGCACTCGTGGGACGGCAACGAGCTGAAGAACCGCGGCTCTGGCGGAGCTGTGT	4819
Qy	4051	CCAG	CACTGACAGTCTACATGGTCAGCTATTATGAGCGGAATCACAAATAGCAGTT	4110
Db	4820	CCT	GCACTGCAAGTTTAAACAGATCTGCACTGCGCCCAACACCGCGCATCGCGTG	4879
Qy	4111	GGAG	CTGCCATGTTTCAGTGGCCCTGTACGACATCCGGAATGCGAAATGTACGACATC	4170
Db	4880	GGCG	CAACAGTGGCAACCTGGCCATTTACGAGCTGGCGCCAGAACAAAGTCCAGATGATC	4939
Qy	4171	CAT	GACACAGGACCAATACATGCACTGGCTTTTCTCTCTGATGGAAGATATCTTGCC	4230
Db	4940	CCG	GCCACAGCATTCGATCACTCGCTGCTCTCTCTGCGAGACCTCAACGGGAGTGTCCGA	4999
Qy	4231	ACCT	ACTCAACACTGACAGCCCAATTTCTTTTGGCAGATGACACGTCATCTGCTGGGA	4290
Db	5000	TCGT	ACTCTTGGCAGAGAACCGGCTCTCTCTCTGCGAGACCTCAACGGGAGTGTCCGA	5059
Qy	4291	AGCA	TGCGCATGTGTAATCTGGGACCTGAGCTGGCTGCTGCAATTAACACTACAGGTGCC	4350
Db	5060	CTGG	CGC-----CAGTCCGACAGCGGCTGTACGAAGGGCTTACTCTACGCGCA	5104
Qy	4351	CCT	GTGAGCCCGGCTCCCGGCTCCACAAATCCCTCAAGCTGGCGCCCGCTCATCTGG	4410
Db	5105	CCCA	TTCGATGTGCTCGC-----GCTTGAATCCGATGCGCCTTGGCAAGAGCTTGTGTGG	5158
Qy	4411	ACT	TCCAAACCGCAACGTCACTCTCATGTGGCCCATGACGCGGAGGAGCACCGCTTCATGGTC	4470
Db	5159	ATCA	CAATCGACCGGTGACCTCATGTGTCGCGACGGCTCCGAGACGCGATTCAAATGTC	5218
Qy	4471	TAA	4473	
Db	5219	TAA	5221	

RESULT 13
ABL19260
ID ABL19260 standard; DNA; 8411 BP.
XX
AC ABL19260;

DT	26-MAR-2002	(first entry)
XX	Drosophila melanogaster	genomic polynucleotide SEQ ID NO 9253.
XX	Drosophila	developmental biology; cell signalling; insecticide;
XX	KW	pharmaceutical; gene; ds.
XX	OS	Drosophila melanogaster.
XX	WO	200171042-A2.
XX	27-SEP-2001.	
XX	23-MAR-2001;	2001WO-US009231.
XX	23-MAR-2000;	2000US-0191637P.
XX	11-JUL-2000;	2000US-00614150.
XX	(PEKE)	PE CORP NY.
XX	Venter JC,	Adams M, Li PWD, Myers EW;
XX	WPI;	2001-656860/75.
XX	New isolated	nucleic acid detection reagent for detecting 1000 or more
XX	genes from	Drosophila and for elucidating cell signaling and cell-cell
XX	interactions.	
XX	Claim 1;	SEQ ID NO 9253; 21pp + Sequence Listing; English.
XX	The invention	relates to an isolated nucleic acid detection reagent
XX	capable of	detecting 1000 or more genes from Drosophila. The invention is
XX	useful in	developmental biology and in elucidating cell signalling and
XX	cell-cell	interactions in higher eukaryotes for the development of
XX	insecticides,	therapeutics and pharmaceutical drugs. The invention
XX	discloses	genomic DNA sequences (ABU16176-ABU30511), expressed DNA
XX	sequences	(ABU1840-ABU16175) and the encoded proteins (ABBS7737-
XX	ABB72072).	The sequence data for this patent did not form part of the
XX	printed	specification, but was obtained in electronic format directly
XX	from	WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	Sequence	8411 BP; 2232 A; 2053 C; 2136 G; 1990 T; 0 U; 0 Other;
XX	Query Match	8.8%; Score 393; DB 4; Length 8411;
XX	Best Local	Similarity 46.6%; Pred. No. 4.4e-107;
XX	Matches	2153; Conservative 0; Mismatches 2275; Indels 195; Gaps 20;
Qy	1	ATGGCAGGAACAGCCTTGTTCTTACCATTTGTTTGGGGTGGAAAGCGCCACACAT 60
Db	2834	ATGGTGGAGTACCAATCTGGTGGTACCCGGTGTCTGGGGTCCACCGCGCCACGCAC 2893
Qy	61	TGCATCTCAGCGGTACTTTTTAAACAGATGATGGGGCCACGATCGTAAACAGGATGTACAGAC 120
Db	2894	TGCATCTTAGCGTGTCTCTCGAGCATCAGTTTACCCTGGTTCACCGGTGTCTACGAT 2953
Qy	121	GGACAAATATGTCTCTGGGATCT---TTCACTAGTACTGCAATTAATCTTCGAGCACTG 177
Db	2954	GGACAGATCTGCCTTTTGGCAAGTGGAGCCACCACCCCTGAAGATGTCCGCGCGGTGCCTC 3013
Qy	178	TTGTTTGGTTCATACAGCATCAATCACATTTGTTTGTCTAAAGCTTGCTTCCAGTGACAAA 237
Db	3014	TTGTTGGGCCATTTCGGCGCCCGTTTGTGCTTAGTGGCTGGCTCGCTTCTGCGGAGNAC 3073
Qy	238	CAGTATATTGTGATGTCATCTGAAAGTGGAGAGATGTGCCCTCTGGGATGTGAGTATGGC 297
Db	3074	AAC TTCCTGCTCAGCTCGTCTGAGAACGGAGAGATGTGCACCTGGGATCTTACCGGAGCGC 3133
Qy	298	AGATGATTTGAATTTACAAATTAGCTTGCACATACTGGCATACAGTCTTACAGTTTC 357
Db	3134	AAGTGCATGAGGCGGTCTCAAGCTTCCACAGGTGCACACCCAAATTCAGAGTACCAACC 3193
Qy	358	TCTGTTGGGAATCAGCGAGAGGAAGGCTTTTATGCCACGACATTTACCTGGAATCCTT 417

Db 3194 -----CCCAACAGCGAGGATGTGGCGCTCTTTTGGCATTGGCTACTATGCTGAAATCATG 3247
Qy 418 GTTGTGATGTACACGCTTGAAGTATATACCTCTTGTAGTATCAAAAGATATACCCAGAC 477
Db 3248 GTCATGATCCGTTACGCTTAGAGTGATCTACGTGCTCAGCTCAAAAGGTCAAAACCGGAT 3307
Qy 478 TGGATTAGCTCCATGATATATTTCGATCCACCGACACACAGAGGACAGTGGTAGCA 537
Db 3308 TGGATATCCGCTATTTACAGTACTGCGCCCATGCGTGCAGAGGATGACGTAGTGTGGCC 3367
Qy 538 CTCTCGGTGATGGCATCTCTGAAGGTCTGGATTGTATTACCTCGGAAATAAAGTGACATCGAG 597
Db 3368 ATCACCACACCGGAACCGTGAAGGTCTGACCTGAC-----GGGTACAGAGACAAA 3421
Qy 598 GATATGAGCAATATTTGAGGAGGAATCAAACCAATTTATTTGTCAGAAATGCCAAAGC 657
Db 3422 CAGCGGAGGCCATTTACAGAGAACGAATCAAAGGAGATCGCTCGCTGAAATGCCATCACC 3481
Qy 658 ATCTCTTTTGTGCAATTTACAAAGGTCACTTTTGGTGTGTTGTTCCAAATATTGGAGG 717
Db 3482 ATGAATGTCTGCGCCAGAACCCAGCGCACAGTGTCTTTAGTTGTACCAATATCTGGCAG 3541
Qy 718 GTGTTCGATCCGAGAGATATTCCTGTGTGTTTCAGGTCTAGTGAATAATGGACAGACA 777
Db 3542 ATCTACGATCCGCGGACTTCATCTGTACTCTGCTCCGTAATGCTCCCGCAAGGAACT 3601
Qy 778 TGGACCGGGGGGACTTTGTCTCATCAGATAAAGTCATCAATTTGGACAGAAAATGGGCAA 837
Db 3602 TGGCAGGGCGGACTTCATCACCTCCGACCGGGTAATGTATGGACAGATGAGGGAAG 3661
Qy 838 AGTTATATTTACAACTACCTGCGAGTTGCTTCCAGCTAGTATTCATTTCCGAGTGAT 897
Db 3662 GGCTACTTTGACAGCTGCCCGCAACTGCAATTTCCGACAAAGGAGTTCCACTCGAAG 3721
Qy 898 GTGGGAAGCGAGTTGAAATTT-----TAATCTCTCTGTACAAACATATCTCTTGGATC 952
Db 3722 AGTGTGTGAGGATGACCTTACCTCTACTAGTGTGCAACAGCGCGGTGATAGGTG 3781
Qy 953 GAAAGATAAAGAGTTGCTAATTTGTCTCTCTGTACTCGGTTCTTCTATGGATGCAAG 1012
Db 3782 CTGTGTCGCCCCCAGCCATGAAGTTGCTTTCAGGGCGCGGTGGCCAGCAATCTACTG 3841
Qy 1013 AATATTTCCATAAATGTTAATTCAGGGTGATTTCTTGAAGGTTGAATTTGGACAA 1072
Db 3842 CGCGGAGACTCCGAGGGATACATCTCTGTCTGGAATGTGCGGAGGTGCGGTGTAGCAAT 3901
Qy 1073 TATCAGACACAGCTGATAAAGAGGAGTGAAGAGGCTGGCAATGACAACT-TCTATT 1131
Db 3902 ATTAGCATCTGACGGCCAAACAGATGCCCGGACCCCTCAAGCGCACGTGTGCACT 3961
Qy 1132 AGTTTGAAGAGGCATTTGATAAATCTGAATCTTGTCTGCTGGAATTTATAGATCAGCTG 1191
Db 3962 TCGTTGTGTCAGGGTGTGTCATTAATGATCCACCCAGTGGGCATCTCTGGATCAGC-- 4019
Qy 1192 AGTGTGATTTCCGAATGATGAACCTCTTAAAGTAACTGCAAGTGTGTTACATCCAGCA 1251
Db 4020 -----TCTCCCGCATACAGAAATCGCGGTGAAGCTCACTCGAGCATCTATCTGCCACAG 4075
Qy 1252 CATGGACGACTGTTTGTGTGTGAGATGGAAGCATAGTTATTGTACCTGCCACACAG 1311
Db 4076 CAGATCGCTGTGTATCGGACCGGAGATGGGACGATAGTCATAGTGGCAGCCACCAG 4135
Qy 1312 ACGGCCATAGTACAGCTGTTTGAAGGGGAACATAGCTCAGAGAGAGGTTGGCCACCTCAC 1371
Db 4136 ACAGTGATGATGAGCTGCTGGTGGGCATCAAGCAAAATCTCAGGATTTGCCCTCGCAT 4195
Qy 1372 AGAACATCTCGTGTGATCGGAACAAAGTCAATGTTTGTCTATATCTCTCATAGGTCTCA 1431
Db 4196 CAGATCTCTACGCTCATCTGTGAAGGGTCAATTTGCCCTGTCTCTGCCCTCGATGATTCAC 4255
Qy 1432 GCTCGGTATGATCAAGATACCTGATATCTGAGGTGTGATTTTTCAGTCAATTTGG 1491
Db 4256 TCGAGATATGAGAAATCGCATCTACTTTTCGCGGGGCATGATTTGCGCGGTGTGTGTGG 4315

Qy 1492 GACATATTTTCTCGAGAAATGAAACATATCTTCTGTGTTTCATGTTGTGTGATTAATCTCAA 1551
Db 4316 GATCTGTACAGCGGAAGCTTTACTGCTAGTGTTTTGGCTTTCATGCGGTGAGATCACACAA 4375
Qy 1552 CTCTAGTTTCCACTCGAAACTCTAGTGTGAAGAGTACAGCACTGCTCTCTGTGAGCC 1611
Db 4376 CTCTGTGTGCGCCAGAGAGCTGTAGTCCAAAGATCTCTGAATGCAATATGTTTGGTGGCC 4435
Qy 1612 AGTGACCACTCAGTAGGACTTTCTAAGTTTTCGAGAGAAAAAATGCAATAATTTGGCACT 1671
Db 4436 TCGGACCAATTCCTTACGCTAGTGTCCCTTTCAGAGCGCAAGTGTGTGACCTTGGCAAGC 4495
Qy 1672 CGTCACTTTTCTTCTATTAAGTAATCAATTAAGAGGCTTCTGATGATTAATCTGTTGGT 1731
Db 4496 AGACATCTATTCTCTGTGTGTACCATTAATGCGTCCCTTGGACGACTTTCTTAATCGTA 4555
Qy 1732 GGTGTGTTTCAGATGTTCTGTGTACGTTGCAATGATGATGATGATGATGATGATGATGAT 1791
Db 4556 GGTGTGTTTCAGATGTTCTGTGTACGTTGCAATGATGATGATGATGATGATGATGATGAT 4615
Qy 1792 GTGATGGGATTAACAGCAGTTGAGATTTCTAAACGCTTGTGATGAA----- 1836
Db 4616 CTTACGGGATGCTGCGAGAGGAAGTCTTTCCGCTTTCGATGAACAGGCGGAGGATGCG 4675
Qy 1837 -GCTGTTCCTGTGCTGTGTGATTTCACTTAGTCTATCGAGAGTCAACCTAAACAGCTAT 1895
Db 4676 GGTCTGCGGGCGCGGTGTTCCAAATGTTGTCATCTGCCAGGAGATGGGTATGCGCAAT 4735
Qy 1896 GACGAGAGTGTCTTCTCTCTTAAATAATGCG-----CCATCATAG 1941
Db 4736 CCGGAGTGCATCTTTCGTTGCGCTGAAAGTCCGCAACATGAATGCAATTCCTGCTCAT 4795
Qy 1942 CTACAAACCTTTCGAACTAACCTCTTGGCTTCTGAGGCACTTCGACAGGGA----- 1994
Db 4796 ACCAGCGGGGATTAGCAACTTCAGCAACTTCAGGTCATTAACAGAGAACTTCGAC 4855
Qy 1995 --TTTACTTAAATTTCTCATAACTCCCTGATGTTCAAGCAATAAAGACAAACCTTAACA 2052
Db 4856 TTTCTGATGAAACACCGCAGCAATCTTTAGTATCCAGGCTCTTCTGTAATAATCTTAAG 4915
Qy 2053 GACCGGACATACATGCTGTATTTCTTGTGTTGAGAGCTTGTATTAATCACTCTGACT 2112
Db 4916 GACCGGAGAGCCACATACCTTTCTTGTACATCGAGGACTGATCTTCGAATTCGATAGC 4975
Qy 2113 GAAGAAAGCTCTAGGCGCAATCTGCTCTTATTTCCCGAGAGAAATTTGCAAAAGCATCT 2172
Db 4976 GAGAAATAGCAGATGACACAGCTACCTTGGATCACTTGGTGTCCACTGCGAAT 5035
Qy 2173 GGCAGTTTACAGAAAGGGGCTCTTTTAACTGGAAGAGCAGCAGCAGTCTTCTTCAA 2232
Db 5036 CCCAAAGATGGCAAAATCGATGCAATTTGGATGCTTAGCAAGAGATCGGTGACTTCTTCAAC 5095
Qy 2233 CAGTCAAGAGAAAGATCAAGAGCAATCAAGCAACCTCTTGTATGATGAGAGGAG 2292
Db 5096 AAGTGAAGAAACAGGCCGTGATGTTGGAAGATCTTGAAGCAATCTTGAAGCAAAACACGGG 5155
Qy 2293 GATGAGAGATATGAGGCGAGAGAGGGA----- 2321
Db 5156 CTGCTGAGAGTTCAAGGGAATACTGAGATTTGTCGAGAGAAAGTACAGGCCAGGTG 5215
Qy 2322 -----AGAAAGTATCTGAAATATCGGTCCAGCAAAATCAAGGCCA 2361
Db 5216 GAGAGCCTGCAAAAGCGAGTAGAGCCGACGAGGAGCAGCAGGACCTGAAAGAGCAAAATC 5275
Qy 2362 TTGACCTTATGAAATATTAATTTAATATGACACATGCAAGAGCTGTTATGCTCTGCTT 2421
Db 5276 GCTTCAAGATGAGGTTACGCAATGTTATGAGAGTAGCCAGCTACTTCTTTCGCTGCTC 5335
Qy 2422 CAGCCTGGGGTTGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAAT 2481
Db 5336 CATTTTGGGGCTTAGATCCATCTAGACAGATGTCGAGAGCGGCTTGGGCTTGTCTT 5395

XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 13190.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX Drosophila melanogaster.
OS WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
PA Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR P-PSDB; ABB62133.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 13190; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 13668 BP; 3652 A; 3342 C; 3582 G; 3092 T; 0 U; 0 Other;
Query Match 8.8%; Score 393; DB 4; Length 13668;
Best Local Similarity 46.6%; Pred. No. 6.2e-107;
Matches 2153; Conservative 0; Mismatches 2275; Indels 195; Gaps 20;
QY 1 ATGGCAGGAAACAGCCCTTGTCTTACCCATTGTTCTTTGGGGTCGAAAGCGCCACACAT 60
DB 3491 ATGGTGAGTACCAATCTGGTGGTACCGGTGCTGTGGGGTCCACCGCGCCACGCAC 3550
QY 61 TGCATCTCAGCGGTACTTTTAAAGATGATGGGGCCAGATCTGTACAGATGTACAGC 120
DB 3551 TGCATCTCAGCGGTACTTTTAAAGATGATGGGGCCAGATCTGTACAGATGTACAGC 3610
QY 121 GGACAAATATGCTCTGGGATCT--TTACAGTAACTGCAAAATTAATCTCCAGCAGCTG 177
DB 3611 GGACAGATCTGCTCTGGGATCTTTGGCAAGTGGAGCCACCACTTGAAGATGTCCGCGGCTC 3670
QY 178 TTGTTTGGTCATACAGCATCAATCACTTTGTTGTTCTTAAAGCTTGTGCTTCCAGTGACAA 237
DB 3671 TTGTTGGCCATTTGGCGCCCGTTTGTGCTTAGTGGTGGCTCGCTTCTGCGGAGAAC 3730
QY 238 CAGTATATGTAGTGCATCTGAAAGTGGAGAGATGGTCCCTCTGGAGATGTAGTATGGC 297
DB 3731 AACTTCCTGCTAGCTGCTCTGAGAACGGAGAGATGTGCACTCTGGGATCTTACGGACGGC 3790
QY 298 AGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 357
DB 3791 AAGTGATGAGGGGCTCAAGCTTCCACAGGTGCACACCCAAATTCAGAGCTACACACC 3850

QY 358 TCTGTTGGGAATCAGCGAGAGAGAGGCTTTTATGCCAGGACATTTACCTGAAATCCTT 417
DB 3851 -----GCCAACAGCGAGGATGTGCGCCTCTTTTGCATTTGGCTACTATGCTGAATCATG 3904
QY 418 GTTGTGATGTCTACAGCCCTTGAAGTATTATATCTCTTAGTATCAAAAGATATCACAGAC 477
DB 3905 GTCATGTGATCCGTTTCAAGCTTAGAGGTGATCTAGTCTAGCTCAAGGTCAAAACCGGAT 3964
QY 478 TGGATTAGTCTCCATGAGTATTATTCGATCCCAACCAACAGAGGACACAGTGGTAGCA 537
DB 3965 TGGATATCCGCTATTTCACGCTACTGCGCCCATCGTCCGAAGGATGACGTAGTGTGGCC 4024
QY 538 CTCTCGTGAATGGGATCTGAAAGTCTGATGTTTACCTCGGAAATAGTGCATGACAG 597
DB 4025 ATCACCAACCAACCGAAACCGTGAAGGTCTGAGCTCTGAC-----GGGTACAGGAAACAA 4078
QY 598 GATACGTAGCCCAATATTTGAGGAGGAATCCAAACCAATTTATTTGTGAGATTTCCCAAAGC 657
DB 4079 CACGCGAGGCCAATTTACGAGAACGAATCCAGGAGATCCGCTGCTGATGCCATCACC 4138
QY 658 ATCTCTTTTGTGCAATTTACAAAGGTCACTTTTGTGTTGTGTTTCCAAATATTGGAGG 717
DB 4139 ATGAATTTGCTGCGCCAGAACCCAGCGCACAGTCTCTTAGTTTGTACCAATATCTGCGAG 4198
QY 718 GTGTTGATGCGCGAGACTATTCTTGTGTTGTGTTAGTCTTAGTGAATAATGACAGACA 777
DB 4199 ATCTACGATGCGCGGCACTTCACTGTACTCTGCTCGTAAATTTCTCCGCAAGGAAACGT 4258
QY 778 TGGACCGGGGGGACTTTGCTCATCAGATAAAGTCAATCATTTGGACAGAAATGGCAA 837
DB 4259 TGGACGGCGGCACTTCACTCCGACCGGGTAAATGCTATGGACAGATAGGGAAG 4318
QY 838 AGTTATATTACAAACTACCTGCGAGTTGCTTCCAGCTAGTGAATTCATTTCCGCAAGTAT 897
DB 4319 GGTACTTTGACAGCTGCCGCAAACTGATTTCCGACAAAGGAGTTCCACTCGAAG 4378
QY 898 GTGGGGAAGCGAGTTGAAATTT-----TAAATCTCTCTGTACAAATATCTCTTGGATC 952
DB 4379 AGTGTGTCAGGAGTGCACCTTACCTCTACTACTGTGTGCAACACGCGGTGATAAGTG 4438
QY 953 GAAAGATAAAGTTGCTAAATTTGCTCTCTGTTACTCGGTCTTCTTATGATGAGAG 1012
DB 4439 CTGCTGCTGCCCGCCAGCCATGAAGTTGCTTTCAGGGCGCGGTGGCCAGCAATCTACTG 4498
QY 1013 AATATTTCCATAAACTGTTAAATTCAGGGTGAATTTCTTCTGGAAGTTGAAATTTTGAACA 1072
DB 4499 CGCGGAGATCTCCGAGGAGTACATCTCTGTCTGGAATGTCCCGGAGGTGCGGTAGACAT 4558
QY 1073 TATCAGACACAGTGTATAAACAGGGAAGTGAAGAAGGGCTGGCAATGACAACT-TCTATT 1131
DB 4559 AATTAGCATCTGCAAGCCAAACAGATGCGCGCGCACCTTCAAGCGCGACGCTGTGCAC 4618
QY 1132 AGTTTGCAGAGGCAATTTGATAACTGAATCTTGTCTGCTGCTGGAATATAGATCAGCTG 1191
DB 4619 TCGTTTGCAGGCGGTGGTCCATTAATGATCCACCAAGTGGGCACTCTTGGATCAGC-- 4676
QY 1192 AGTGTGATTTCCAAATAGTAACTGAACTCTTAAAGTAACTGCAAGTGTGTACATACAGCA 1251
DB 4677 -----TCTCCGATCACAGATTCGCGGTGAGAGTCACTCTCGAGCATCTATCTGCCACAG 4732
QY 1252 CATGACGACTTTGTTGCTGCTGGAAGATGGAAGCATAGTTATTTGCTGCTGCCACACAG 1311
DB 4733 CAGAGTCCGCTGCTGATCGGACGAGGATGGCAGCATAGTCTAGTGTGCGAGCACCCAG 4792
QY 1312 ACGGCATAGTACAGCTGTTGCAAGGGAACACATGCTCAGAGAGGTTGGCCACCTCAC 1371
DB 4793 ACAGTGATGATGAGCTGCTGTTGGGCACTCAAGCAAACTTTCAGCGATTTGGCCCTCGCAT 4852
QY 1372 AGAACACTCCGTTGCTCATCGGAACAAAGTCAATGTTTGTCTATATCTCTCATCAGGTCTCA 1431
DB 4853 CAGATCTCTACGCTCATCTGGAAGGGTCAATTTGCTGCTCTGCGCCCTCGATGATTCAC 4912
QY 1432 GCTCGGTATGATCAAGATACCTGATATCTGGAGGTGTGGATTTTTTTCAGTCAATATTTGG 1491

PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241122P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.

PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-476159/51.
DR P-PSDB; ADM20106.
XX Isolated nucleic acid molecule encoding a channel/transporter protein is
PT used in preventing, treating or ameliorating a medical condition.
XX Claim 1; SEQ ID NO 434; 809pp; English.

XX The invention relates to an isolated nucleic acid molecule encoding a
CC channel/transporter protein or sequences at least 95% identical to a
CC these. The nucleic acids and proteins encoded by them are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. The antibodies to the proteins can also be used
CC in alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiodenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC The polypeptides can also be used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC also be used as a food additive or preservative to increase or decrease
CC storage capabilities. This sequence corresponds to a gene of the
CC invention.

XX Sequence 491 BP; 119 A; 131 C; 118 G; 121 T; 0 U; 2 Other;

Query Match 4.3%; Score 191.8; DB 5; Length 491;
Best Local Similarity 94.2%; Pred. No. 9.1e-47;
Matches 210; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 1 ATGGCAGGAAACAGCCTTGTCTTACCCATGTCTTTGGGTGGAAGGCCACACAT 60
|||||
Db 220 ATGGCAGGAAACAGCCTTGTCTTACCCATGTCTTTGGGTGGAAGGCCACACAT 279
QY 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACCATCGTAAACAGGATGTACGAC 120
|||||
Db 280 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACCATCGTAAACAGGATGTACGAC 339
QY 121 GGACAAATATGTCTCTGGGATCTTTTCAGTAGAATCTTCAAAATTAATCTCGAGCACTGTTG 180
|||||
Db 340 GGACAAATATGTCTCTGGGATCTTTTCAGTAGAATCTTCAAAATTAATCTCGAGCACTGTTG 399
QY 181 TTGGTTCATACAGCATCAATCACTTG- TTTGCTAAAGCTTGT 222
|||||
Db 400 TTGGTTCATACAGCATCAATCACTTGTTTGTCTAAAGCTTGT 442

Search completed: July 2, 2005, 04:15:25
Job time : 2170 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 00:59:56 ; Search time 18357 Seconds
(without alignments)
11806.950 Million cell updates/sec

Title: US-10-645-335-1
Perfect score: 4473
Sequence: 1 atggcaggaaacagcttgt.....agcacgcttcagtgtctaa 4473

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.cm.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4473	100.0	4473	9	AY099325 Homo sapi
2	4469.8	99.9	4548	6	AX244640 Sequence
3	4258.8	95.2	4371	9	AF188125 Homo sapi
4	3549.8	79.4	4497	6	AX244642 Sequence
5	3546.8	79.3	4467	10	AF188124
6	3337.8	79.1	4467	10	AF105813
7	3337.8	79.1	4494	6	AX244638
8	3497	78.2	6224	6	AR541843
9	3485	77.9	6072	9	AB011113
10	3382.8	75.6	4371	10	AF192379
11	3369.2	75.3	4371	10	AF188123
12	2807.4	62.8	3316	6	AX714633
13	2807.4	62.8	3316	9	AK057141
14	2680	59.9	5905	10	AK122304
15	1841.8	41.2	4460	9	BC050352
16	1788.8	40.0	3154	5	BC073055
17	1164	26.0	1436	6	CQ720875
18	969.8	21.7	1916	6	CQ719779
19	773	17.3	164076	9	AC008006

C	20	773	17.3	189371	2	AC090230	AC090230 Homo sapi
C	21	613	13.7	183544	10	AC102132	AC102132 Mus muscu
C	22	603.4	13.5	215898	2	AC096428	AC096428 Rattus no
C	23	491	11.0	6104	9	BSM808720	BSM808720 Homo sapi
C	24	487.4	10.9	5900	9	BSM805990	BSM805990 Homo sapi
C	25	474.6	10.6	2578	6	AX833927	AX833927 Sequence
C	26	474.6	10.6	2578	9	AK096055	AK096055 Homo sapi
C	27	395.2	8.8	48287	2	AC017774	AC017774 Drosophil
C	28	393	8.8	5221	6	CQ598374	CQ598374 Sequence
C	29	393	8.8	6049	3	BT015183	BT015183 Drosophil
C	30	393	8.8	8411	6	CQ598373	CQ598373 Sequence
C	31	393	8.8	13668	6	CQ578837	CQ578837 Sequence
C	32	393	8.8	169618	3	AC105055	AC105055 Drosophil
C	33	393	8.8	170490	3	AC104347	AC104347 Drosophil
C	34	393	8.8	300933	3	AE003422	AE003422 Drosophil
C	35	382.8	8.6	29352	3	DMC8684	DMC8684 Drosophil
C	36	293	6.6	72383	2	AC101173	AC101173 Mus muscu
C	37	287.2	6.4	5292	9	BSM808902	BSM808902 Homo sapi
C	38	253	5.7	171185	2	AP002424	AP002424 Homo sapi
C	39	253	5.7	171747	2	AC087678	AC087678 Homo sapi
C	40	253	5.7	177097	2	AP001569	AP001569 Homo sapi
C	41	253	5.7	179726	9	AC007052	AC007052 Homo sapi
C	42	253	5.7	182411	2	AC090408	AC090408 Homo sapi
C	43	226.8	5.1	211466	9	AC012301	AC012301 Homo sapi
C	44	218.2	4.9	2046	3	AY051974	AY051974 Drosophil
C	45	208.2	4.7	250880	10	AC102040	AC102040 Mus muscu

ALIGNMENTS

RESULT 1
AY099325
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
CDS

AY099325
Homo sapiens rabconnectin-3 beta mRNA, complete cds.
4473 bp mRNA linear PRI 10-JUL-2003
AY099325
Homo sapiens rabconnectin-3 beta mRNA, complete cds.
4473 bp mRNA linear PRI 10-JUL-2003
AY099325.1
GI:30313408
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 4473)
Kawabe,H., Sakisaka,T., Yasumi,M., Shingai,T., Izumi,G., Nagano,F.,
Deguchi-Tawarada,M., Takeuchi,M., Nakanishi,H. and Takai,Y.
A novel rabconnectin-3-binding protein that directly binds a
GDP/GTP exchange protein for Rab3A small G protein implicated in
Ca(2+)-dependent exocytosis of neurotransmitter
Genes Cells 8 (6), 537-546 (2003)
22671656
12786944
2 (bases 1 to 4473)
Kawabe,H., Nakanishi,H., Shingai,T. and Takai,Y.
Direct Submission
Submitted (23-APR-2002) Biochemistry and Molecular Biology, Osaka
University, Yamada-Oka 2-2, Suita, Osaka 565-0871, Japan
Location/Qualifiers
1. .4473
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1. .4473
/note="brain enriched protein"
/product="rabconnectin-3 beta"
/protein_id="AA033134.1"
/translation="MAGNSLVLPVLMGRKAPHCISAVLLTDDGATIVTGDHGDIC
LWDLGVBLQINFRALLFGHTASITCLSKACASDQYIVSASESMCLWVSDGRCI
EPTKLACTHTGTFQFVGNOREGRLLCHGHPILVVDATSLVSLVSKISPDW
ISSMSGIIRSHRTOEDTVVALSVTGILKQVITVSEISDMQDTEPIFEESKPIYCNQCQ
SISFCAPQSRLLVCSKYNRPVFDAGDGLSCSPSENGQTWGTGDFVSSDKVILWTE
NQSQYIYKLPASDPSFRSDVSKAVENLIPPVQHILLDRKDKELLCPPTTRFFY

GCREYFKLLIQDSSRLNIWNI SDTADKQSEGLAMTTSISLQEA FDKLPNCPAG
IIDLQSVIPNSNEPLKVTASVYIPAHGRVLCVREDGSIIVIPATQTAIVOLLQGEHML
RWGPHRRTLRHNRKVCCLLPYHQVSARYDORYLISGVDFSVIIMDLFSEMHKPIF
CVHGSEITQLLVPPENC SARVQHCISVADSHVGLLSUREKKIWLADIPSEMHKPIF
KWRPSDYLVGCGSDGVVWQMDTGALDRCVNGITAVEILNACDEADPAVPAVLSHP
AVNLKQAMTRRSILAALKNMAHKLQTLATNLLASEASDKGNLPKYSHNSLMVQAIKTN
LTPDQTHVLFPDVEALIIOLLTEASRPNTALISPENLOKASGSDSKGGSFLTGKRAA
VLPOQVETIKENIKHEHLLDREDEEINRORRESDPEYRSKSKPLTLLVNLVMTD
TAKLFMSCLHAWGLNVLDEVCLDRGLMKLPCHTQSVFGLLSRGHMSLMLPGYNQAC
KLSHGKTEVGRKLPASGEVKGTYGVSRVATVTOHLLSIIISLANTLSMTNATFIDGHH
KKGKTRPPSPPTDLSKARGSPPTSNIVQQTQKQVAPVVSARSADNSGDPSPAS
ALHTCFLENGSQLAAMHCWMLPDILGLDKFRPILLEMARRWODRCLEVEAREAQV
LLAELRRIEQAQRKEAIDAWAPVLPQYIDHVISPGVTSAAOTITTPADSPAERAAQV
EEHDLVDDIITGCLSSVPQMKKISTSYEERKQATAVLLGVIGAEFGAIEBPKL
LTKPRSSQYPEGGLTSGSNYSLARHTCKALTELLQPPSPKLPHPHSIRRTALIDL
IGRFTVWPEPMDVSAVLMGLLELCADAEKQANITMGLPLSPADASARSARLSLIL
ATAPRPAFTITTAKEVHRHTALAAANTQSQONMHTTTLARAKEBIEILRVIEBILKEMPTD
VVDLLVEMWDIIIMYCLEGSLVKKGLQCFPAICRFYVMVSYVERNHRTAVGARHGSVA
LYDIRTKCQOTIGHKGPITAVAFAPDGRYLATYSNTDSHISFWMQNTSLLSIGMLN
SAPQLRACIKTYQVPPVPASPQSHNALKLARLLWTSNRNVILMAHDGKEHRFMV"

ORIGIN

Query Match 100.0%; Score 4473; DB 9; Length 4473;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 4473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGCGAGGAAACAGCCCTTGTTCTACCCGATGTTCTTTGGGTCGAAAGCGGCCACACAT	60
DB	1	ATGCGAGGAAACAGCCCTTGTTCTACCCGATGTTCTTTGGGTCGAAAGCGGCCACACAT	60
QY	61	TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAAACAGATGTCACGAC	120
DB	61	TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAAACAGATGTCACGAC	120
QY	121	GGCAAAATATGTCCTCGGATCTTTTCAGTAGAACTGCAAAATTAATCCTCGAGCACTGTTG	180
DB	121	GGCAAAATATGTCCTCGGATCTTTTCAGTAGAACTGCAAAATTAATCCTCGAGCACTGTTG	180
QY	181	TTTGGTCATACAGCATCAATCACTTGTTGTCTTAAAGCTTGCTTCCTCGAGTGACAAACAG	240
DB	181	TTTGGTCATACAGCATCAATCACTTGTTGTCTTAAAGCTTGCTTCCTCGAGTGACAAACAG	240
QY	241	TATATTGTAGTGCATCTGAAAGTGGAGATGTCCTCTGGAGTGTGAGTGATGCGACA	300
DB	241	TATATTGTAGTGCATCTGAAAGTGGAGATGTCCTCTGGAGTGTGAGTGATGCGACA	300
QY	301	TGTATTGAATTTACAAATTTAGCTTGCACATACATCTGGCATACAGTTCTACCAAGTTCTCT	360
DB	301	TGTATTGAATTTACAAATTTAGCTTGCACATACATCTGGCATACAGTTCTACCAAGTTCTCT	360
QY	361	GTTGGGNAATCAGCGAGAGGAGGCTTTTATGCGACGGACNTTACCTGNAATCCTTGTT	420
DB	361	GTTGGGNAATCAGCGAGAGGAGGCTTTTATGCGACGGACNTTACCTGNAATCCTTGTT	420
QY	421	GTGAGTCTCAGCGCTTGAAGTATTATATCTCTTAGTATCAAGATATACCCAGACTGG	480
DB	421	GTGAGTCTCAGCGCTTGAAGTATTATATCTCTTAGTATCAAGATATACCCAGACTGG	480
QY	481	ATTAGTCTCCATGAGTATTATTTCGATCCCAACGACACAGAGGACAGTGGTAGCACTC	540
DB	481	ATTAGTCTCCATGAGTATTATTTCGATCCCAACGACACAGAGGACAGTGGTAGCACTC	540
QY	541	TCGGTGATCGGCATCTGTAAGGTCGTGATTGTTTACCTCGGAAATAAGTGACATGCAAGAT	600
DB	541	TCGGTGATCGGCATCTGTAAGGTCGTGATTGTTTACCTCGGAAATAAGTGACATGCAAGAT	600
QY	601	ACTGAGCCATATTTTCAGGAGGAATCCAAACCAATTTATTGTGAGAAATGCCAAAGCATC	660
DB	601	ACTGAGCCATATTTTCAGGAGGAATCCAAACCAATTTATTGTGAGAAATGCCAAAGCATC	660
QY	661	TCCTTTTGTGCAATTTACAAAGGTCACTTTTGGTGTGTGTTTCCAAATATTGCGAGGTG	720
DB	661	TCCTTTTGTGCAATTTACAAAGGTCACTTTTGGTGTGTGTTTCCAAATATTGCGAGGTG	720

QY	721	TTTCGATCGCGAGACTATTTCCTTGTTGTGTTTCAGGTCTCTAGTGAAAAATGACAGACATGG	780
DB	721	TTTCGATCGCGAGACTATTTCCTTGTTGTGTTTCAGGTCTCTAGTGAAAAATGACAGACATGG	780
QY	781	ACCGGGGGGAGCTTTGTCTCATCAGATAAAGTCATCATTTTGACAGAAAAATGGGCAAGT	840
DB	781	ACCGGGGGGAGCTTTGTCTCATCAGATAAAGTCATCATTTTGACAGAAAAATGGGCAAGT	840
QY	841	TATATTTTACAACTACCTCCAGTTCCTCTCTGTAACAATATCTCTTGGATCGAAAGAT	900
DB	841	TATATTTTACAACTACCTCCAGTTCCTCTCTGTAACAATATCTCTTGGATCGAAAGAT	900
QY	901	GGGAGGCGAGTTGAAAAATTTAAATCTCTCTGTAACAATATCTCTTGGATCGAAAGAT	960
DB	901	GGGAGGCGAGTTGAAAAATTTAAATCTCTCTGTAACAATATCTCTTGGATCGAAAGAT	960
QY	961	AAAGAGTTGCTAATTTGTCCTCTCTGTTACTCGGTTCTTCTATGAGTACGAGAAATATTTTC	1020
DB	961	AAAGAGTTGCTAATTTGTCCTCTCTGTTACTCGGTTCTTCTATGAGTACGAGAAATATTTTC	1020
QY	1021	CATAAATCTGTTAATTTACGGGTGATTTCTTCTGGAAGGTTGAATATTTTGGAAACATATCAGAC	1080
DB	1021	CATAAATCTGTTAATTTACGGGTGATTTCTTCTGGAAGGTTGAATATTTTGGAAACATATCAGAC	1080
QY	1081	ACAGCTGATAAACAAGGAGTGAAGAGGCTGGCAATGACAACTTCTATTTAGTTTGCACAA	1140
DB	1081	ACAGCTGATAAACAAGGAGTGAAGAGGCTGGCAATGACAACTTCTATTTAGTTTGCACAA	1140
QY	1141	GAGGCATTTGATAAATACTGTAATCTTGTCTCTGCTGGAATTTATAGATCAGCTGAGTGTGAT	1200
DB	1141	GAGGCATTTGATAAATACTGTAATCTTGTCTCTGCTGGAATTTATAGATCAGCTGAGTGTGAT	1200
QY	1201	CCCAATAGTATAAAGCTCTTAAAGTAACTGCAAGTGTGTATACATACCAGCAATGGACCA	1260
DB	1201	CCCAATAGTATAAAGCTCTTAAAGTAACTGCAAGTGTGTATACATACCAGCAATGGACCA	1260
QY	1261	CTTGTGTTGTCGTCGTAAGTGAAGTGAAGTATTTGTACCTGCCACACAGACGGCCATA	1320
DB	1261	CTTGTGTTGTCGTCGTAAGTGAAGTGAAGTATTTGTACCTGCCACACAGACGGCCATA	1320
QY	1321	GTACAGCTGTTGCAAGGGGAAACATGCTCAGAAGAGGTTGGCCACCTCACAGAACATCTC	1380
DB	1321	GTACAGCTGTTGCAAGGGGAAACATGCTCAGAAGAGGTTGGCCACCTCACAGAACATCTC	1380
QY	1381	CGTGTGTCATCGGAAACAAAGTCAATGTTGCTATATCTCTCATCAGTCTCAGCTCGGTAT	1440
DB	1381	CGTGTGTCATCGGAAACAAAGTCAATGTTGCTATATCTCTCATCAGTCTCAGCTCGGTAT	1440
QY	1441	GATCAAGATACCTGATATCTGAGGTGTGGAATTTTTCAGTCATATAATTTGGGACATATTT	1500
DB	1441	GATCAAGATACCTGATATCTGAGGTGTGGAATTTTTCAGTCATATAATTTGGGACATATTT	1500
QY	1501	TCGGAAGAAATGAAACATATCTTCTGTGTTCAATGTTGTTGAGATTAATCACTCTCTAGTT	1560
DB	1501	TCGGAAGAAATGAAACATATCTTCTGTGTTCAATGTTGTTGAGATTAATCACTCTCTAGTT	1560
QY	1561	CCACCTGAAACCTGATAGTCAAGAGTACAGCACTGCTGCTGCTGTGTAGCCAGTACCCAC	1620
DB	1561	CCACCTGAAACCTGATAGTCAAGAGTACAGCACTGCTGCTGCTGTGTAGCCAGTACCCAC	1620
QY	1621	TCAGTAGGACTTCTAAGTTTTCGAGAGAAAAATGCATAATTTTGGCATCTCTCGTCACTTT	1680
DB	1621	TCAGTAGGACTTCTAAGTTTTCGAGAGAAAAATGCATAATTTTGGCATCTCTCGTCACTTT	1680
QY	1681	TTTCTTATTCAAGTAATCAAAATGGAGGCTTCTCTGATGATTAATCTCGTGTGGGGTGTCA	1740
DB	1681	TTTCTTATTCAAGTAATCAAAATGGAGGCTTCTCTGATGATTAATCTCGTGTGGGGTGTCA	1740
QY	1741	GATGGTCTGTGTAAGTCTGGCAAAATGAGTACCTGTTGCAATTTGGATCGTTGTGTGATGGG	1800
DB	1741	GATGGTCTGTGTAAGTCTGGCAAAATGAGTACCTGTTGCAATTTGGATCGTTGTGTGATGGG	1800

QY 1801 ATRACAGCAGTTGAGATTCTAAACGCTTGATGAAGCTGTTCTGCTGCTGTTGATTCA 1860
DB 1801 ATACAGCAGTTGAGATTCTAAACGCTTGATGAAGCTGTTCTGCTGCTGTTGATTCA 1860
QY 1861 CTTAGTCATCCAGCAGTCAACCTTAAACAAAGCTATGACGAGACGTAGTCTTGCTGCTCTT 1920
DB 1861 CTTAGTCATCCAGCAGTCAACCTTAAACAAAGCTATGACGAGACGTAGTCTTGCTGCTCTT 1920
QY 1921 AAAAATATGGCCCATATAGCTACAAACCCCTTGCAACTAACTCTTGGCTCTTGAGGCA 1980
DB 1921 AAAAATATGGCCCATATAGCTACAAACCCCTTGCAACTAACTCTTGGCTCTTGAGGCA 1980
QY 1981 TCTGACAGGGAATTTTACCTAAATATCTCATAACTCCCTGATGGTTCAAGCAATAAG 2040
DB 1981 TCTGACAGGGAATTTTACCTAAATATCTCATAACTCCCTGATGGTTCAAGCAATAAG 2040
QY 2041 ACAAACTTAACAGACCCGGAATACATATGCTATTCTTTGATGGGAAGGTTGATTAT 2100
DB 2041 ACAAACTTAACAGACCCGGAATACATATGCTATTCTTTGATGGGAAGGTTGATTAT 2100
QY 2101 CAATCTCTGATGAAGAGCTCTAGGCCGAATACTGCTCTTATTTCCCCAGAGAAATTG 2160
DB 2101 CAATCTCTGATGAAGAGCTCTAGGCCGAATACTGCTCTTATTTCCCCAGAGAAATTG 2160
QY 2161 CAAAAGCATCTGCGAGTTGAGCAAAAGGGGCTCTTTTAACTGGAAACGAGCAGCA 2220
DB 2161 CAAAAGCATCTGCGAGTTGAGCAAAAGGGGCTCTTTTAACTGGAAACGAGCAGCA 2220
QY 2221 GTTCTCTTCCAAAGTGAAGAAACGATCAAGAGAAACATCAAGGAACACCTCTTGAT 2280
DB 2221 GTTCTCTTCCAAAGTGAAGAAACGATCAAGAGAAACATCAAGGAACACCTCTTGAT 2280
QY 2281 GATGAAGAGGAGTGAAGAGATAATGAGGCAGAGAGGGAAGGATGATCCTGAATAT 2340
DB 2281 GATGAAGAGGAGTGAAGAGATAATGAGGCAGAGAGGGAAGGATGATCCTGAATAT 2340
QY 2341 CGGTCCAGCAATCAAGACCATGACCCCTATTAGAAATATATTTAACTATGGAACCTGCA 2400
DB 2341 CGGTCCAGCAATCAAGACCATGACCCCTATTAGAAATATATTTAACTATGGAACCTGCA 2400
QY 2401 AAGCTGTTATGCTGCTCTTCAAGCTGCGGCTTGAATGAAGTACTGGAATGAATTTGC 2460
DB 2401 AAGCTGTTATGCTGCTCTTCAAGCTGCGGCTTGAATGAAGTACTGGAATGAATTTGC 2460
QY 2461 CTGGATCGCTTGGAAATGCTGAAACCCCACTGCAACCGTATCGTTGGCTCTTTGTCAAGA 2520
DB 2461 CTGGATCGCTTGGAAATGCTGAAACCCCACTGCAACCGTATCGTTGGCTCTTTGTCAAGA 2520
QY 2521 GGAGGCCATATGTCATGATGCTGCGGGTTAATCAGCGCTGTTGTAACCTGTCAAT 2580
DB 2521 GGAGGCCATATGTCATGATGCTGCGGGTTAATCAGCGCTGTTGTAACCTGTCAAT 2580
QY 2581 GGGAAACAGAGTGAAGAGAGCTGCGAGCTGAGGGAGTGAAGAGGGAACCTTAC 2640
DB 2581 GGGAAACAGAGTGAAGAGAGCTGCGAGCTGAGGGAGTGAAGAGGGAACCTTAC 2640
QY 2641 GGAGTGTCCGTGCGCTGACCAACAGCATCTCTGTCTATCATTTCTTTGGCAAAATACT 2700
DB 2641 GGAGTGTCCGTGCGCTGACCAACAGCATCTCTGTCTATCATTTCTTTGGCAAAATACT 2700
QY 2701 TTAATGATATGACCAATGCAATTTTATTTGGTGATCATATGAAGAGGCTCTACAGG 2760
DB 2701 TTAATGATATGACCAATGCAATTTTATTTGGTGATCATATGAAGAGGCTCTACAGG 2760
QY 2761 CCACCTAGACCAAGCAACCCAGACCTTCTAAGGCAAGGGTTCCCTCCAACTTCCAGT 2820
DB 2761 CCACCTAGACCAAGCAACCCAGACCTTCTAAGGCAAGGGTTCCCTCCAACTTCCAGT 2820
QY 2821 AATATTGTGCAAGCAGAGATTAAACAAAGTTGCTGCACTCTGCTGCTCCGCTCGGTCTGAT 2880
DB 2821 AATATTGTGCAAGCAGAGATTAAACAAAGTTGCTGCACTCTGCTGCTCCGCTCGGTCTGAT 2880
QY 2881 GCTGATCACTCTGGCTCTGACCCCTCTCTGCTCTCTGCTTTTACATACCTGTTTCTTAGTA 2940

DB 2881 GCTGATCACTCTGGCTCTGACCCCTCTCTGCTCTCTGCTTTTACATACCTGTTTCTTAGTA 2940
QY 2941 AATGAAGTTGGAGTCAGTTAGCTGCTATGCACTGTGTATGTCTGCGAGACCTACTTGGGA 3000
DB 2941 AATGAAGTTGGAGTCAGTTAGCTGCTATGCACTGTGTATGTCTGCGAGACCTACTTGGGA 3000
QY 3001 TTGGATAAATTTAGGCTCTCCCTTTCTGGAGATGCTGGCCCGAAGATGCGAAGATCGATGC 3060
DB 3001 TTGGATAAATTTAGGCTCTCCCTTTCTGGAGATGCTGGCCCGAAGATGCGAAGATCGATGC 3060
QY 3061 TTGGAGGTGAGAGAGCGCGCAGAGGCTCTGCTTTCTGGCGAACTGAGAGAAATTTGAGCAG 3120
DB 3061 TTGGAGGTGAGAGAGCGCGCAGAGGCTCTGCTTTCTGGCGAACTGAGAGAAATTTGAGCAG 3120
QY 3121 GCAGGAGGAGAGGAGCGCATTTGATGCTGGGCTCTTACTTACTCTAGTACATAGACAC 3180
DB 3121 GCAGGAGGAGAGGAGCGCATTTGATGCTGGGCTCTTACTTACTCTAGTACATAGACAC 3180
QY 3181 GTCATATCACTGGAGTCAATCAGAAAGCGCGCAGATATACACCGGCTCTCTGATGCC 3240
DB 3181 GTCATATCACTGGAGTCAATCAGAAAGCGCGCAGATATACACCGGCTCTCTGATGCC 3240
QY 3241 TCAGGCGCTGAGCAAAAGTCCAGGAGGAGGAGCATGACCTTTGTCAGATGATCATCACC 3300
DB 3241 TCAGGCGCTGAGCAAAAGTCCAGGAGGAGGAGCATGACCTTTGTCAGATGATCATCACC 3300
QY 3301 ACTGGTTGCTTATCAAGTGTCCCAAAATGAAAAAATTTCTACATCTTACGAGGAAAGA 3360
DB 3301 ACTGGTTGCTTATCAAGTGTCCCAAAATGAAAAAATTTCTACATCTTACGAGGAAAGA 3360
QY 3361 CGGAACCAAGCTACCGCTATTGTTTACTTGGAGTAAATAGGAGCTGAAATTTGGTGTGAA 3420
DB 3361 CGGAACCAAGCTACCGCTATTGTTTACTTGGAGTAAATAGGAGCTGAAATTTGGTGTGAA 3420
QY 3421 ATTGAACCTCTTAACTATGACCAACCTCGAGCTCTAGCCAAATTTCTGAGGATTC 3480
DB 3421 ATTGAACCTCTTAACTATGACCAACCTCGAGCTCTAGCCAAATTTCTGAGGATTC 3480
QY 3481 GGGTTCACTAGTGTGGATCCAACTACTCGTGGCGAGACATATTTGCAAGGACATGACG 3540
DB 3481 GGGTTCACTAGTGTGGATCCAACTACTCGTGGCGAGACATATTTGCAAGGACATGACG 3540
QY 3541 TTTCTTCTGCTACAGCCTCCAAAGCCCAAACTTCTCCACACAGCATATCCGAAGAACA 3600
DB 3541 TTTCTTCTGCTACAGCCTCCAAAGCCCAAACTTCTCCACACAGCATATCCGAAGAACA 3600
QY 3601 GCCATTGATCTGATTTGGAGCTGGGTTTCACTGTTTGGGAGGCTTACATGATGTCGGCT 3660
DB 3601 GCCATTGATCTGATTTGGAGCTGGGTTTCACTGTTTGGGAGGCTTACATGATGTCGGCT 3660
QY 3661 GTTCTGATGGGCTTCTCGAACTTTTGTGCGATGCCGAGAAACAACTTGCCAAACATCACA 3720
DB 3661 GTTCTGATGGGCTTCTCGAACTTTTGTGCGATGCCGAGAAACAACTTGCCAAACATCACA 3720
QY 3721 ATGGGGTTGCTCTGAGCCCGAGCAGCTGACTCGGCCGCTCTGCGAGGATGCGCTCTCG 3780
DB 3721 ATGGGGTTGCTCTGAGCCCGAGCAGCTGACTCGGCCGCTCTGCGAGGATGCGCTCTCG 3780
QY 3781 CTCATTTGCCACCGCAGACCAACCGGCTTTCATCAACCATATAGCCAAAGAGGTACACAGA 3840
DB 3781 CTCATTTGCCACCGCAGACCAACCGGCTTTCATCAACCATATAGCCAAAGAGGTACACAGA 3840
QY 3841 CATAGGCTCTTGGAGCAAAATACCCCAATCAGAGCAATATGCACAACTCTTTGCA 3900
DB 3841 CATAGGCTCTTGGAGCAAAATACCCCAATCAGAGCAATATGCACAACTCTTTGCA 3900
QY 3901 CGAGCTAAAGGGGAAATTTTGGAGGTCATTGAAATTTCTTATTTGAAAAGATGCCACAGAT 3960
DB 3901 CGAGCTAAAGGGGAAATTTTGGAGGTCATTGAAATTTCTTATTTGAAAAGATGCCACAGAT 3960
QY 3961 GTTGTGATCTTCTGCTGGAGGTTATGGACATCATTTATGCTACTGCTTGAAGATCTTTA 4020

Db 3961 GTTGTGGATCTTCTCGTGAGGTTATGAGACATCATTTATGTACTGCTTGAAGGATCTTTA 4020
Qy 4021 GTTAAAAAGAAAGTCTTCAAGATGTTTCCAGCCATCTGACAGTCTTACATGTCAGC 4080
Db 4021 GTTAAAAAGAAAGTCTTCAAGATGTTTCCAGCCATCTGACAGTCTTACATGTCAGC 4080
Qy 4081 TATTATGAGCGGAATCACAGAATAGCAGTTGGAGCTCGCATGTTTCAGTGGCCCTGTAC 4140
Db 4081 TATTATGAGCGGAATCACAGAATAGCAGTTGGAGCTCGCATGTTTCAGTGGCCCTGTAC 4140
Qy 4141 GACATCGGACATGGAATATGTCAGACATTCATGAGACACAGGACCAATCACTGCGAGT 4200
Db 4141 GACATCGGACATGGAATATGTCAGACATTCATGAGACACAGGACCAATCACTGCGAGT 4200
Qy 4201 GCTTTTGTCTCTGATGGAAGATATCTTGCACCTACTCAAACTGACGACCCATTTCT 4260
Db 4201 GCTTTTGTCTCTGATGGAAGATATCTTGCACCTACTCAAACTGACGACCCATTTCT 4260
Qy 4261 TTTTGGCAGATGAACAGTCACTGCTGGAAGCATCGGATGCTGAACCTGGCACCTCAG 4320
Db 4261 TTTTGGCAGATGAACAGTCACTGCTGGAAGCATCGGATGCTGAACCTGGCACCTCAG 4320
Qy 4321 CTGCGTGCATTAACACCTACAGGTGCCCTGTGACGCGCGCTCCCGGCTCCAC 4380
Db 4321 CTGCGTGCATTAACACCTACAGGTGCCCTGTGACGCGCGCTCCCGGCTCCAC 4380
Qy 4381 AATGCCCTCAAGCTGCGCGGCTCATCTGACCTTCAACCGCAACGTCATCTCATGGCC 4440
Db 4381 AATGCCCTCAAGCTGCGCGGCTCATCTGACCTTCAACCGCAACGTCATCTCATGGCC 4440
Qy 4441 CATGACGGGAAGGAGCAGCGCTTCATGCTCTAA 4473
Db 4441 CATGACGGGAAGGAGCAGCGCTTCATGCTCTAA 4473

RESULT 2

AX244640
LOCUS AX244640 4548 bp DNA linear PAT 28-SEP-2001
DEFINITION Sequence 3 from Patent WO0166739.
ACCESSION AX244640
VERSION AX244640.1 GI:15859534
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Thorpeitson,S.S. and Sanders,S.
[dna encoding a tag gene (tgf-_g(b) resistance associated gene)
and its protein product
Patent: WO 0166739-A 3 13-SEP-2001;
THE DEPARTMENT OF HEALTH & HUMAN SERVICES (US)
Location/Qualifiers
1. .4548
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
76. .4548
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC8575.1"
/db_xref="GI:15859535"
/translation="MAGNSLVLPDIVWGRKAPTHCISAVLLTDDGATVITGCHDQIC
LWDSVLEQVNPRLALFGHTASITCLSKACASDQKQIVSASESGMCLWDVSDGRCI
EFTKLACTHGIQYQFVSQNGRGLLGHGYEILVVDATSLVLYSLVSKISPDW
ISSMSIIRHRTQEDTVALSVILKWLVIWSEISDMODTPIFEESKPIYQNCQW
SISFCAFTORSLLVVCSSKYRVPDAGDYLSCSPSENGTWTGDFVSSDKVIIWTE
NGQSYIKLPASCLPASDSFRSDVGAENLPPVQHILLDRKDKELLICPPVTFPY
GCREYFKLLIQDSSGLRNINISDTADKQSGEGLMATTISLQEARDKLNPFCAG
IIDLQSVIPNSPEKLTASVYIPAHGRVLCGRDGSIVIPATQTAIVOLLQGEHL
RRGWPHRTLRHNRKVTCLLYPHQVSARYDQRYLSGGVDFSVIIVDFSGMKHIF
CVHGGEITOLLVPPENCARVQHCISVADSHSVGLLSLRKKCI MLASRHLPIQVI
KWRPSDDLYLVGCGSDGSVYVWQMDTGALDRVCVMGITAVEILNACDEAVPAADVSLSHP

FEATURES

source
1. .4548
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
76. .4548
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC8575.1"
/db_xref="GI:15859535"

CDS

1
Thorpeitson,S.S. and Sanders,S.
[dna encoding a tag gene (tgf-_g(b) resistance associated gene)
and its protein product
Patent: WO 0166739-A 3 13-SEP-2001;
THE DEPARTMENT OF HEALTH & HUMAN SERVICES (US)
Location/Qualifiers
1. .4548
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
76. .4548
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC8575.1"
/db_xref="GI:15859535"
/translation="MAGNSLVLPDIVWGRKAPTHCISAVLLTDDGATVITGCHDQIC
LWDSVLEQVNPRLALFGHTASITCLSKACASDQKQIVSASESGMCLWDVSDGRCI
EFTKLACTHGIQYQFVSQNGRGLLGHGYEILVVDATSLVLYSLVSKISPDW
ISSMSIIRHRTQEDTVALSVILKWLVIWSEISDMODTPIFEESKPIYQNCQW
SISFCAFTORSLLVVCSSKYRVPDAGDYLSCSPSENGTWTGDFVSSDKVIIWTE
NGQSYIKLPASCLPASDSFRSDVGAENLPPVQHILLDRKDKELLICPPVTFPY
GCREYFKLLIQDSSGLRNINISDTADKQSGEGLMATTISLQEARDKLNPFCAG
IIDLQSVIPNSPEKLTASVYIPAHGRVLCGRDGSIVIPATQTAIVOLLQGEHL
RRGWPHRTLRHNRKVTCLLYPHQVSARYDQRYLSGGVDFSVIIVDFSGMKHIF
CVHGGEITOLLVPPENCARVQHCISVADSHSVGLLSLRKKCI MLASRHLPIQVI
KWRPSDDLYLVGCGSDGSVYVWQMDTGALDRVCVMGITAVEILNACDEAVPAADVSLSHP

ORIGIN

Query Match 99.9%; Score 4469.8; DB 6; Length 4548;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4471; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGGCAGGAAACAGCGCTTGTCTTACCCATTTGTTTGGGGTGGAAAGCGCCACACAT 60
Db 76 ATGGCAGGAAACAGCGCTTGTCTTACCCATTTGTTTGGGGTGGAAAGCGCCACACAT 135
Qy 61 TGCATCTCAGCGGTACTTTTACAGATGATGGGGCCAGATCGTAACAGAGTGTCCAGAC 120
Db 136 TGCATCTCAGCGGTACTTTTACAGATGATGGGGCCAGATCGTAACAGAGTGTCCAGAC 195
Qy 121 GGACAAATATGTCTCTGGGATCTTTTCACTAGTAAGTGCAGAAATTAATCTCTGAGCAGTGT 180
Db 196 GGACAAATATGTCTCTGGGATCTTTTCACTAGTAAGTGCAGAAATTAATCTCTGAGCAGTGT 255
Qy 181 TTTGGTCATACAGATCAATCACTTGTCTTAAAGCTTGTGCTTCCAGTGCACAAACAG 240
Db 256 TTTGGTCATACAGATCAATCACTTGTCTTAAAGCTTGTGCTTCCAGTGCACAAACAG 315
Qy 241 TATATTGTGAGTGCATCTGAAAGTGGAGAGTGTGCTCTGGGATGTGAGTGTATGGCAGA 300
Db 316 TATATTGTGAGTGCATCTGAAAGTGGAGAGTGTGCTCTGGGATGTGAGTGTATGGCAGA 375
Qy 301 TGTATTGAATTTACAAATTTAGCTTTGCACACATATCGGCATACAGTGTTCACAGTTCCT 360
Db 376 TGTATTGAATTTACAAATTTAGCTTTGCACACATATCGGCATACAGTGTTCACAGTTCCT 435
Qy 361 GTTGGGAATCAGCGAGAGGAGGCTTTTATGCCAGGACATTAACCTGAAATCTTGT 420
Db 436 GTTGGGAATCAGCGAGAGGAGGCTTTTATGCCAGGACATTAACCTGAAATCTTGT 495
Qy 421 GTGGATGCTACCGAGCTTGAAGTATTATCTCTTAGTATCAAGATATCACAGACTGG 480
Db 496 GTGGATGCTACCGAGCTTGAAGTATTATCTCTTAGTATCAAGATATCACAGACTGG 555
Qy 481 ATTAGTCCATGAGTATTATTCGATCCCAACCGAACACAGAGGACACAGTGTAGCAGTCT 540
Db 556 ATTAGTCCATGAGTATTATTCGATCCCAACCGAACACAGAGGACACAGTGTAGCAGTCT 615
Qy 541 TCGGTGACTGGCATCTGAGGCTTGGATTTTACCTCGGAATTAAGTGCACATGCAAGAT 600
Db 616 TCGGTGACTGGCATCTGAGGCTTGGATTTTACCTCGGAATTAAGTGCACATGCAAGAT 675
Qy 601 ACTGAGCCAAATATTTGAGGAGGAATCCAAACCAATTTATTTGTGAGAAATGCCAAGCATC 660
Db 676 ACTGAGCCAAATATTTGAGGAGGAATCCAAACCAATTTATTTGAGAAATGCCAAGCATC 735
Qy 661 TCTTTTGTGCAATTTACACAAAGGTCACITTTTGGTGTGTGTTCCTTCCAAATATTTGGAGG 720
Db 736 TCTTTTGTGCAATTTACACAAAGGTCACITTTTGGTGTGTGTTCCTTCCAAATATTTGGAGG 795
Qy 721 TTCGATGCCGAGACATTTCTTGTGTGTTCAGTCTCTAGTGCCTTCTGAGAAATGCCAAGCATC 780
Db 796 TTCGATGCCGAGACATTTCTTGTGTGTTCAGTCTCTAGTGCCTTCTGAGAAATGCCAAGCATC 855

ANVLMQAMTRRSALAAKNMAHHKLTQTLATNLIASEASDKGNLPKYSHNSLMVQAIKTN
LTPDPIHVLFPDVBALIIQLTEASRENTALISPENLQKASGSDKGGSFUTGKRA
VLPOQKETIKENIKHILLDEDEEETIMORRESDPEYRSSKSKPLTLRYNLTM
TAKFMSCLHAWGLNEVLDEVLCLRLMKLPHCTVSGLLSRGHMSLMPLGYNPAC
KLSHGKTVEGRKLPASGEGVGKTYGVSRVTTQHLLSIISLANTLMSMTNATFIDHGM
KKGPTPRPSTPDLSKARGSPPTSSNVQOIKQVAPVVSADSNHSGDPPSPAS
LLHFCFLVNEGMSQLAAMHCVMLPDLGLDKPPLLEMLRWRDRCLLEVREAQAL
ALABURRTEOAGRKEAIDAWAPYLPQYIDHVTSPGVTSBAQOTITAPDAPPEAKVO
EEEDLDVDDITTCGLSSVPQMKKISTSYERRKOATAILVLGVIGAFEGARIEPKL
LTPRSSQIEPEGFLTGGSGNSYLRHTKALITLLOPPSPKLPPISTIRTAIDL
ITGRFTWEPYNDVSAVLMGELLECADAEKQALNTMGLPLSPADARSARSAHLSLI
LADPPAPITTIKEVHRHTALANTQSQNMHTTTLARAKEIILRVTEILTERKPTD
VVDLLVEVMDIIMYCLEGLVKKGLOCFPAICRFYMVSYVERHNRITAVGARHGSVA
LYDIRTKCOTIHKHGPITAVAPADGRYLATYNTSDSHISFWOMNTSLIGSGLN
SAPQLRCIKTYQVPPVQSPASPGSHNALKLARLIWTSNENVILMAHDGKEHRPMV"

Db 3016 AATGAAGTTGGAGTCAGTTAGCTGTGCTATGCACTGTGTTATGCTGCGACAGCTACTGGGA 3075
Qy 3001 TTGGATAAATTTAGGCTCCCTTCTGGAGATGCTGGCCGGAAGATGGCAAGATCCATGC 3060
Db 3076 TTGGATAAATTTAGGCTCCCTTCTGGAGATGCTGGCCGGAAGATGGCAAGATCCATGC 3135
Qy 3061 TTGGAGGTGAGAGAGCCGACAGGCGCTGCTTCTGGCGAAGCTGAGAGAGATTTGAGCAG 3120
Db 3136 TTGGAGGTGAGAGAGCCGACAGGCGCTGCTTCTGGCGAAGCTGAGAGAGATTTGAGCAG 3195
Qy 3121 GCAGGCGAGGAAGCAATGTATGCTGGCTCTTACTTACTCTAGTACATAGACAC 3180
Db 3196 GCAGGCGAGGAAGCAATGTATGCTGGCTCTTACTTACTCTAGTACATAGACAC 3255
Qy 3181 GTCATATCACTGGAGTCACATCAGAAGCCGCGACATATCAGCAGGCTCTGATGCC 3240
Db 3256 GTCATATCACTGGAGTCACATCAGAAGCCGCGACATATCAGCAGGCTCTGATGCC 3315
Qy 3241 TCAGGGCTGAAGCAAAAGTCCAGGAGGAAGCATGACCTTGTGACGATGACATCACC 3300
Db 3316 TCAGGGCTGAAGCAAAAGTCCAGGAGGAAGCATGACCTTGTGACGATGACATCACC 3375
Qy 3301 ACTGGTTGCTTATCAAGTGTCCACAAATGAAAAAATTTCTACATCTTACGAGGAAGA 3360
Db 3376 ACTGGTTGCTTATCAAGTGTCCACAAATGAAAAAATTTCTACATCTTACGAGGAAGA 3435
Qy 3361 CGAAGCAAGCTACCGCTATTTGTTTACTTGGAGTAATAGGAGCTGAAATTTGGTGTGAA 3420
Db 3436 CGAAGCAAGCTACCGCTATTTGTTTACTTGGAGTAATAGGAGCTGAAATTTGGTGTGAA 3495
Qy 3421 ATTGAACCTCTAAATCTATTGACACAGACCTCGAAGCTTAGCCAAATTTCTGAGGATTC 3480
Db 3496 ATTGAACCTCTAAATCTATTGACACAGACCTCGAAGCTTAGCCAAATTTCTGAGGATTC 3555
Qy 3481 GGGTTGACTAGTGGTGATCAACTACTCGCTGGCCAGACATATTTGCAAGGCACTGACG 3540
Db 3556 GGGTTGACTAGTGGTGATCAACTACTCGCTGGCCAGACATATTTGCAAGGCACTGACG 3615
Qy 3541 TTTCTTCTGCTACAGCTTCAAGCCCAACTTCTCCACACAGCACTATCCGAAAGCA 3600
Db 3616 TTTCTTCTGCTACAGCTTCAAGCCCAACTTCTCCACACAGCACTATCCGAAAGCA 3675
Qy 3601 GCATTTGATCTGATTTGAGCTGGGTTTCACTGTTTGGAGCTTACATGATGTCGCT 3660
Db 3676 GCATTTGATCTGATTTGAGCTGGGTTTCACTGTTTGGAGCTTACATGATGTCGCT 3735
Qy 3661 GTTCTGATGGGCTTCTCGAACTTTTGGCGATGCGGAGAAACAACTTGCCAAATCACA 3720
Db 3736 GTTCTGATGGGCTTCTCGAACTTTTGGCGATGCGGAGAAACAACTTGCCAAATCACA 3795
Qy 3721 ATGGGGTTGCTCTGAGCCACAGCACTGACTCGGCCGCTCTCGAGGCATGCCCTCTCG 3780
Db 3796 ATGGGGTTGCTCTGAGCCACAGCACTGACTCGGCCGCTCTCGAGGCATGCCCTCTCG 3855
Qy 3781 CTATTTGCCACCGCAGACCAACCGCTTCTATCACCACCACTAGCCAAAGAGGTACACAGA 3840
Db 3856 CTATTTGCCACCGCAGACCAACCGCTTCTATCACCACCACTAGCCAAAGAGGTACACAGA 3915
Qy 3841 CATACGCTCTTGACGAAATACCCAAATCAAGCAGATATGACACAACTTTGCA 3900
Db 3916 CATACGCTCTTGACGAAATACCCAAATCAAGCAGATATGACACAACTTTGCA 3975
Qy 3901 CGAGCTAAGGGGAAATTTGAGAGTCATTGAAATTTCTTATGAAAGATGCCACAGAT 3960
Db 3976 CGAGCTAAGGGGAAATTTTGGAGAGTCATTGAAATTTCTTATGAAAGATGCCACAGAT 4035
Qy 3961 GTTGTGATCTTCTGTTGAGGTTTATGGACATCAATATGACTGCTTGAAGGATCTTTA 4020
Db 4036 GTTGTGATCTTCTGTTGAGGTTTATGGACATCAATATGACTGCTTGAAGGATCTTTA 4095
Qy 4021 GTTAAAAAGAAAGTCTTCAAGATGTTTCCAGCCATCTGCAAGGTTCTACATGGTCAGC 4080
Db 4096 GTTAAAAAGAAAGTCTTCAAGATGTTTCCAGCCATCTGCAAGGTTCTACATGGTCAGC 4155

Qy 4081 TATTATGAGCGAATCACAGAATAGCAGTTGGAGCTCGCCATGTTCACTGGCCCTGTAC 4140
Db 4156 TATTATGAGCGAATCACAGAATAGCAGTTGGAGCTCGCCATGTTCACTGGCCCTGTAC 4215
Qy 4141 GACATCCGGACTGGAAAAATGTGAGCAAAATCCATGGACACAAGGGACCAATCACTGCAGTG 4200
Db 4216 GACATCCGGACTGGAAAAATGTGAGCAAAATCCATGGACACAAGGGACCAATCACTGCAGTG 4275
Qy 4201 GCTTTTGTCTCTGATGGAAGATATCTTGCCACTACTCAAACTGACAGCCACATTTCT 4260
Db 4276 GCTTTTGTCTCTGATGGAAGATATCTTGCCACTACTCAAACTGACAGCCACATTTCT 4335
Qy 4261 TTTTGGCAGATCAACACGTCACCTGCTGGGAAGCATCGGCATCTGAACTCGGCACCTCAG 4320
Db 4336 TTTTGGCAGATCAACACGTCACCTGCTGGGAAGCATCGGCATCTGAACTCGGCACCTCAG 4395
Qy 4321 CTGCGCTGCATTTAAACCTTACCAGGTGCCCCCTGTGCGAGCCCGCTGCCCGGCTCCAC 4380
Db 4396 CTGCGCTGCATTTAAACCTTACCAGGTGCCCCCTGTGCGAGCCCGCTGCCCGGCTCCAC 4455
Qy 4381 AATGCCCTCAAGCTGCCCCGGCTCATCTGGACTTCCAAACCGCAACGTCATCTCATGGCC 4440
Db 4456 AATGCCCTCAAGCTGCCCCGGCTCATCTGGACTTCCAAACCGCAACGTCATCTCATGGCC 4515
Qy 4441 CATGACGGAAGGAGCACCGCTTCATGGTCTAA 4473
Db 4516 CATGACGGAAGGAGCACCGCTTCATGGTCTAA 4548

RESULT 3

AF188125 4371 bp mRNA linear PRI 15-SEP-2001
LOCUS Homo sapiens TGF-beta resistance-associated protein TRAG (TRAG)
DEFINITION mRNA, partial cds.
ACCESSION AF188125
VERSION AF188125.1 GI:15624074
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 4371)
AUTHORS Sanders, S. and Thorgeirsson, S.S.
TITLE TRAG: a novel gene associated with TGF-beta resistance
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4371)
AUTHORS Sanders, S. and Thorgeirsson, S.S.
TITLE Direct Submission
JOURNAL Submitted (20-SEP-1999) LEC, NIH, 37 Convent Drive, Bldg. 37, Rm 3C28, Bethesda, MD 20892-4255, USA
FEATURES
source
1. .4371
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21.1"
/tissue_type="liver"
1. .>4371
/gene="TRAG"
1. .>4371
/gene="TRAG"
/codon_start=1
/product="TGF-beta resistance-associated protein TRAG"
/protein_id="AA03983.1"
/translation="MAGNSLVLPVILWGRKAPTHICISAVLLTDDGATVITGCHDQIC
LWDLVELQVNPRLPLFGHTASITCLSKACASSDKQYIVSASESEMCILWVSDGRCI
ETFKLACTHTGTFQFVSIGNOREGRLLCHGHPYELLVVDATSLVLYSLVSKI
ISMSIIRSHRTQEDTVVALSVGILKWIWTSBISDMQDTEPFIEESKPIYQNCQ
STSPCAFTQSVLLVVCYSRFRVDFDGLSCGSPSENGQTWTGDFVSSDKVIWTE
NQSQSYIKLPASCLPASDSFRSDVGKAVENLIPPVQHILLDRKDELLICPPVTREFFY

GCREVPHKLLIQDSSRLNINISDTADKQSEGLAMTTSISLQEAEPDKLNPCCPAG
IIDLQSVIPNSPELKVATSVYIPAGRLVQCBEDSGSIVIVATQAIYVOLLQGHMFL
RRPPHRTLGRNKNVTCLLYPHQVSARYDQRYLISGGVDFSVIWIIDIPSGEMKHF
CVHGSIITOLLVPENCARSARVHCICSVASDHSVGLLSREKXKICMLASRLHFPPIQVH
KWRPSDDYLVVCGSDSVVWQMDTGALDRVCVMTAVELINACDEAVPAADVLSLHP
AVNLKQAMTRRSLLAKNMAHKLQTLATNLLASEASDKGNLPKYSHNSLWQAIKTN
LTPDPIHLVFPDVEALIQLLTERASRPNTALISPENLOKASGSDKGSPLTGKRAA
VLFOQVKETIKENIKHLLDDEEDEINRQREESDPYRSKSKPLFLLYNLTMD
TKLGFMSCLHAGNLNEVLDECLDRGLMKPHCTVSFGLLSRGGHMSLMPGYNQPA
KLKSHGTVEYRKLPAESEGKGTGYVSRAVTTQHLLSII SLANTLMSMTNFIIDHM
KKGPTRPSPDLKARGSPPTSSNI VQGOIKQWSQLAAHCVMLPDLGLGDKFR
PPLLEMLARWODRCLVREAAQALLAELRIEORAGKEAIDAWAPVOYIDIVIS
PGVTSAAOTIITAPDASPEAKVQSEEDLVDDDITTCCLSSVPOMKKSISYSEER
KQATAYVLVGLGAEFGASIEBPPLKLTTRPSSQIPPEGGLTSGGSNYSLARHTKAL
TELLQGPSPKUPPHSTIRRTADLIGRTFVWEPYMDVSAVIMGLMELCADAEKOLA
NITWGLPSPADARSARHLSLIATAPPAFTTIAKEVHRHTLAAANTQOQNPH
TTTLARAKGEILIRVIBILI EKMPDVIDLLVEVMDIIMYLEGSLVKKGLQECFPPI
CRPYMYSYERNHRIAVGARHGSVALYDRTGKCTIYHGHKGPITAVAPAPGRVLT
YSTDSHISFWMQNTSLGSI GMLNSAPQLRCIKTVQVPVPPQSPGSHNALKLARLI
WTSNRNVILMAHDGKHERFMV"

ORIGIN

Query Match 95.2%; Score 4258.8; DB 9; Length 4371;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 4369; Conservative 0; Mismatches 2; Indels 99; Gaps 1;
QY 1 ATGGCAGGAAACAGCCTTGTTTACCCATGTTCTTTGGGGTGGAAAGCGCCACACAT 60
DB 1 ATGGCAGGAAACAGCCTTGTTTACCCATGTTCTTTGGGGTGGAAAGCGCCACACAT 60
QY 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAAACAGGATGTACGAC 120
DB 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGGCCACGATCGTAAACAGGATGTACGAC 120
QY 121 GCACAAATATGTCCTCGGGATCTTTCAGTAGAACCTGCAAAATAATCTCTGAGCAGCTGTTG 180
DB 121 GCACAAATATGTCCTCGGGATCTTTCAGTAGAACCTGCAAAATAATCTCTGAGCAGCTGTTG 180
QY 181 TTTGGTCATACAGCATCAATCACTGTTGTTGTCTTAAAGCTTGTGCTTCCAGTGACAAACAG 240
DB 181 TTTGGTCATACAGCATCAATCACTGTTGTTGTCTTAAAGCTTGTGCTTCCAGTGACAAACAG 240
QY 241 TATATTGTGAGTGATCTGAAAGTGAGAGATGTGCTCTGGGATGTGAGTATGTCAGAG 300
DB 241 TATATTGTGAGTGATCTGAAAGTGAGAGATGTGCTCTGGGATGTGAGTATGTCAGAG 300
QY 301 TGTATTGAATTTACAAAATTTAGCTTCACACATCTGCGCATACAGTTCTTACCAGTTCTCT 360
DB 301 TGTATTGAATTTACAAAATTTAGCTTCACACATCTGCGCATACAGTTCTTACCAGTTCTCT 360
QY 361 GTTGGGAATCAGCGAAGGAAGGCTTTTATGCGACGACATTAACCTCGAAATCTCTTGT 420
DB 361 GTTGGGAATCAGCGAAGGAAGGCTTTTATGCGACGACATTAACCTCGAAATCTCTTGT 420
QY 421 GTGGAATCTACAGCCTTGAAGTATTAATCTCTTAGTATCAAGATATCAACAGACTGG 480
DB 421 GTGGAATCTACAGCCTTGAAGTATTAATCTCTTAGTATCAAGATATCAACAGACTGG 480
QY 481 ATTAGCTCCATGATTAATTTAGTCCACACATCTGCGCATACAGTTCTTACCAGTTCTCT 540
DB 481 ATTAGCTCCATGATTAATTTAGTCCACACATCTGCGCATACAGTTCTTACCAGTTCTCT 540
QY 541 TCGGTGACTGGCATCTGAGGCTTGGATTGTTTACCTCGGAATAAGTACATGACAGGAT 600
DB 541 TCGGTGACTGGCATCTGAGGCTTGGATTGTTTACCTCGGAATAAGTACATGACAGGAT 600
QY 601 ACTGAGCCAAATTTGAGGAGGAATCCAAACCAATTTATTTGTCAGAAATGCCAAGCATC 660
DB 601 ACTGAGCCAAATTTGAGGAGGAATCCAAACCAATTTATTTGTCAGAAATGCCAAGCATC 660
QY 661 TCTTTTGTGCAATTTACAAAGGTCATTTTGGTGTGTGTTGTTCCAAATATTGGAGGTG 720
DB 661 TCTTTTGTGCAATTTACAAAGGTCATTTTGGTGTGTGTTGTTCCAAATATTGGAGGTG 720

QY 721 TTCGATGCCGAGAGACTATTCCTTGTGTGTTTGTTCAGGTCCTAGTGAAAAATGACAGACATGG 780
DB 721 TTCGATGCCGAGAGACTATTCCTTGTGTGTTTGTTCAGGTCCTAGTGAAAAATGACAGACATGG 780
QY 781 ACCGGGGGGGACTTTTGTCTCATCAGATAAAGTCATCATTTGGACAGAAAAATGGGCAAAAGT 840
DB 781 ACCGGGGGGGACTTTTGTCTCATCAGATAAAGTCATCATTTGGACAGAAAAATGGGCAAAAGT 840
QY 841 TATATTTACAAACTACTCCGAGTTGCTTCCAGCTAGTGAATTCATTCGCGCAGTGATGTG 900
DB 841 TATATTTACAAACTACTCCGAGTTGCTTCCAGCTAGTGAATTCATTCGCGCAGTGATGTG 900
QY 901 GGGAGGCGAGTTGAAAAATTTAATTCCTCTCTGTACAAACATATCTCTTTGGATCGAAAAGAT 960
DB 901 GGGAGGCGAGTTGAAAAATTTAATTCCTCTCTGTACAAACATATCTCTTTGGATCGAAAAGAT 960
QY 961 AAAGAGTTGCTAATTTGTCT 1020
DB 961 AAAGAGTTGCTAATTTGTCT 1020
QY 1021 CATAAACTGTTAATTCAGGGTGAATTCCTCTGGAAGGTTGAAATATTTGGAAACATATCAGAC 1080
DB 1021 CATAAACTGTTAATTCAGGGTGAATTCCTCTGGAAGGTTGAAATATTTGGAAACATATCAGAC 1080
QY 1081 ACAGCTGATAAACAGGGAAGTGAAGAGGGCTGGCAATGACAACTTCTATTATTAGTTGCAA 1140
DB 1081 ACAGCTGATAAACAGGGAAGTGAAGAGGGCTGGCAATGACAACTTCTATTATTAGTTGCAA 1140
QY 1141 GAGGCATTTGATAAATGAACTGAATCCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
DB 1141 GAGGCATTTGATAAATGAACTGAATCCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
QY 1201 CCCAATAGTAATAAGAACTCTTAAAGTAAGTGAAGTGTCTACATACAGACACATGACACGA 1260
DB 1201 CCCAATAGTAATAAGAACTCTTAAAGTAAGTGAAGTGTCTACATACAGACACATGACACGA 1260
QY 1261 CTTGTTTGTGGTCTGTAAGATGGAAGCATAGTTATTGTACCTGCGCACACAGACGGCCATA 1320
DB 1261 CTTGTTTGTGGTCTGTAAGATGGAAGCATAGTTATTGTACCTGCGCACACAGACGGCCATA 1320
QY 1321 GTACAGCTGTTTGCAGAGGGGAAACATGCTCAGAGAGGTTGGCCACCTCAGACACACATC 1380
DB 1321 GTACAGCTGTTTGCAGAGGGGAAACATGCTCAGAGAGGTTGGCCACCTCAGACACACATC 1380
QY 1381 CGTGTGTCATCGGAAACAAAGTCACATGTTGTGTATATCTCTCATCAGCTCTCAGCTCGGTAT 1440
DB 1381 CGTGTGTCATCGGAAACAAAGTCACATGTTGTGTATATCTCTCATCAGCTCTCAGCTCGGTAT 1440
QY 1441 GATCAAGATACCTGATATCTGGAGGTGTGGAATTTTTCAGTCATATTTTGGGACATATTTT 1500
DB 1441 GATCAAGATACCTGATATCTGGAGGTGTGGAATTTTTCAGTCATATTTTGGGACATATTTT 1500
QY 1501 TCTGAGAAAATGAAAACATATCTTCTGTGTTTCAATGTGTGTGAGATTTACTCAACTCTAGTT 1560
DB 1501 TCTGAGAAAATGAAAACATATCTTCTGTGTTTCAATGTGTGTGAGATTTACTCAACTCTAGTT 1560
QY 1561 CCACCTGAAAACCTGTAGTGCAGAGTACAGACTGCTCATCTGCTGTAGCCAGTACACAC 1620
DB 1561 CCACCTGAAAACCTGTAGTGCAGAGTACAGACTGCTCATCTGCTGTAGCCAGTACACAC 1620
QY 1621 TCAGTAGGACTTCTAAGTTTTCGAGAGAAAAATGCAATAATGTTGGCATCTCTGTCACCTT 1680
DB 1621 TCAGTAGGACTTCTAAGTTTTCGAGAGAAAAATGCAATAATGTTGGCATCTCTGTCACCTT 1680
QY 1681 TTTCTATTCAAGTAATCAAAATGAGGCTTCTGATGATTAATCTGGTGTGGGGTGTTC 1740
DB 1681 TTTCTATTCAAGTAATCAAAATGAGGCTTCTGATGATTAATCTGGTGTGGGGTGTTC 1740
QY 1741 GATGTTCTGTGTCAGTCTGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1800
DB 1741 GATGTTCTGTGTCAGTCTGGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1800

Qy	1801	ATAACAGCAGTGTGAGATTCTTAAACGCTTGTGATGAGAGCTGTTCTGCTGCTGTTGATTCTCA	1860
Db	1801	ATAACAGCAGTGTGAGATTCTTAAACGCTTGTGATGAGAGCTGTTCTGCTGCTGTTGATTCTCA	1860
Qy	1861	CTTTAGTTCATCCAGCAGTCAACCTTAAACAAGCTATGACGAGAGTGATGCTTTGCTGCTCTT	1920
Db	1861	CTTTAGTTCATCCAGCAGTCAACCTTAAACAAGCTATGACGAGAGTGATGCTTTGCTGCTCTT	1920
Qy	1921	AAAAATATGGCCCATCATAAAGCTACAAACCCCTTGGCAACTAAACCTCTTTGGCTTCTTGAGGCA	1980
Db	1921	AAAAATATGGCCCATCATAAAGCTACAAACCCCTTGGCAACTAAACCTCTTTGGCTTCTTGAGGCA	1980
Qy	1981	TCTGCAAGGGAAATTTACTCTAAATATTCCTATCAATCTCCCTGATGGTTCAAGCAATAAAG	2040
Db	1981	TCTGCAAGGGAAATTTACTCTAAATATTCCTATCAATCTCCCTGATGGTTCAAGCAATAAAG	2040
Qy	2041	ACAAACCTTAACAGACCCGGACATACATGCTGCTATTTCTTCATCTGGGAAGGTTGATTATT	2100
Db	2041	ACAAACCTTAACAGACCCGGACATACATGCTGCTATTTCTTCATCTGGGAAGGTTGATTATT	2100
Qy	2101	CAACTCTCTGACTGAAGGAAGCCTCTAGGCGCCGAATACTGCTCTTATTTTCCCCCAGAGAAATTTG	2160
Db	2101	CAACTCTCTGACTGAAGGAAGCCTCTAGGCGCCGAATACTGCTCTTATTTTCCCCCAGAGAAATTTG	2160
Qy	2161	CAAAAAGCATCTGGCAGTTTCAGACAAAGGGGGCTCTTTTTTAACTGGAATAACGAGCAGCA	2220
Db	2161	CAAAAAGCATCTGGCAGTTTCAGACAAAGGGGGCTCTTTTTTAACTGGAATAACGAGCAGCA	2220
Qy	2221	GTTCTCTTCCNAAGGTGAAAGAAACGATCAAGAGAGAACATCAAGNACACCTCCTTGAT	2280
Db	2221	GTTCTCTTCCNAAGGTGAAAGAAACGATCAAGAGAGAACATCAAGNACACCTCCTTGAT	2280
Qy	2281	GATCAAGAGGAGGATGAGGAGATAATCAGGCAGAGAGGGAAGAAAGTGAATCTCTGAATAT	2340
Db	2281	GATCAAGAGGAGGATGAGGAGATAATCAGGCAGAGAGGGAAGAAAGTGAATCTCTGAATAT	2340
Qy	2341	CGGTCCAGCAAAATCAAAGCCATTCGACCTTATTTAGAAATAAATTTAACTATGGAACACTGCA	2400
Db	2341	CGGTCCAGCAAAATCAAAGCCATTCGACCTTATTTAGAAATAAATTTAACTATGGAACACTGCA	2400
Qy	2401	AAGCTGTTTATGTCCTGCTTCAGCCTGGGGTTTGAATGAAGTACTGGAATGAAGTTTC	2460
Db	2401	AAGCTGTTTATGTCCTGCTTCAGCCTGGGGTTTGAATGAAGTACTGGAATGAAGTTTC	2460
Qy	2461	CTGGATCGCCTTGGAATGCTGAAACCCCACTGCACCCGTATCTGTTGGCCCTCTTGTCAAGA	2520
Db	2461	CTGGATCGCCTTGGAATGCTGAAACCCCACTGCACCCGTATCTGTTGGCCCTCTTGTCAAGA	2520
Qy	2521	GGAGGCCATATGTCATGATGCTGCGGGTTTATAATCAGCCTGCTGTTGAAACTGTCACAT	2580
Db	2521	GGAGGCCATATGTCATGATGCTGCGGGTTTATAATCAGCCTGCTGTTGAAACTGTCACAT	2580
Qy	2581	GGGAAAAACAGAAGTAGAAAGGAAGCTGCGACGCTCTGAGGGGTAGGAAAGGGAACCTTAC	2640
Db	2581	GGGAAAAACAGAAGTAGAAAGGAAGCTGCGACGCTCTGAGGGGTAGGAAAGGGAACCTTAC	2640
Qy	2641	GGAGTGTCCTGTCCTGCTCACCACACAGCATCTCTGCTCTATCATTTCTTTGGCAAAATACT	2700
Db	2641	GGAGTGTCCTGTCCTGCTCACCACACAGCATCTCTGCTCTATCATTTCTTTGGCAAAATACT	2700
Qy	2701	TTAATGAGTATGACCAATGCAACTTTTATTTGGTGATCATATGAAGAAGGGTCTTACCAGG	2760
Db	2701	TTAATGAGTATGACCAATGCAACTTTTATTTGGTGATCATATGAAGAAGGGTCTTACCAGG	2760
Qy	2761	CCACCTTAGACCAAGCAACCCAGACCTTTCTAAGCAAGGGGTTCCCTCTCCAACTTCAGT	2820
Db	2761	CCACCTTAGACCAAGCAACCCAGACCTTTCTAAGCAAGGGGTTCCCTCTCCAACTTCAGT	2820
Qy	2821	AATATTGTGCAAGGACAGATTAAACAAGTTGCTGCACTCTGCTTTCCGCTCGGCTCTGAT	2880
Db	2821	AATATTGTGCAAGGACAGATTAAACAAGTTGCTGCTTTCCGCTCGGCTCTGAT	2880
Qy	2881	GCTGATCACTCTGGCTCTGACCCCTCTTCTGCTCTCTGCTTTTACATCACTGTTTCTTAGTA	2940

QY 781 ACCGGGGGACCTTTGTCTCATGATAAGTATCATTTGGACAGAAAATGGCAAGT 840
DB 808 ACTGGAGGGGACCTTTGTGTGTCGAGACAAAGTCATCATCTGACAGAAAATGGCAGAGT 867
QY 841 TATATTTACAAACTACCTGCCAGTTGCCCTTCAGCTAGTGTGATTCATTCCGCGAGTGTG 900
DB 868 TACATCTACAAACTCCCTGCCAGTTGCCCTTCAGCTAGTGTGATTCATTCCGCGAGCGAGTGTG 927
QY 901 GGGAGGCGAGTTGAAATTTAATTCCTCTCTGTACAAATATCTCTTGGATCGAAAGAT 960
DB 928 GGGAAAGCAGTGGAAAATTTGATTCCTCTCTGTGCGAGCATAGCTCTTGGATCAGAAGAT 987
QY 961 AAGAGTTGCTAATTTGCTCTCTCTGTGTTACTCGGTTCTCTATGAGTACAGAGATTTTC 1020
DB 988 AAGAGTTGCTAATTTGCTCTCTCTGTGTTACTCGGTTCTCTACGATGCAAGAAATTTTG 1047
QY 1021 CATAAACTGTAAATTCAGGGGTGATTCCTCTGGAAGGTTGAAATTTTGGAAACATATCAGAC 1080
DB 1048 CATAGCTACTAATTCAGGGCGATTCCTCTGGAAGGTTAAATATCTGGAACATAGCAGAC 1107
QY 1081 ACAGCTGATAAACAGGGAGTGAAGAGGGCTGGCAATGACAACTTCTATTAGTTTGCNA 1140
DB 1108 ATAGCAGAGAAAACAGAGACCGATGAAGGGCTAAAGATGACAACTTGTATTAGCTTGCNA 1167
QY 1141 GAGGCATTTGATAAAGCTGAATCCTCTCTCTGCTGCTGGAATTTATAGATCAGCTGAGTGTGAT 1200
DB 1168 GAGGCATTTGACAAAGCTGAAGCCTTGCCCTGCTGGAATTTATCGATCAGCTAAGTGTGAT 1227
QY 1201 CCCAATPAGTAAGAACTCTTAAAGTAACTGCAAGTGTGTACATACAGCACATGAGACA 1260
DB 1228 CCCAACAGCAATGAACCTCTTAAAGTAACTGCGAGTGTCTACATACAGCACACAGGGCGA 1287
QY 1261 CTTGTTTGTGCTGTAAGATGAAGCATAGTTATTTGTACTGCGCAACAGACAGCGGCATA 1320
DB 1288 CTTGTTTGTGCGCGGAGATGAAGCATCATTTTGTGCTGCGCACCCAGACAGCGGCATA 1347
QY 1321 GTACGCTGTGCAAGGGAAACACATGCTCAGAAAGGTTGGCACCTCACAGAACACATC 1380
DB 1348 GTTCAGCTATTGCAAGGAAACACATGCTCAGAAAGGTTGGCGCCCTCACAGAACCCCTC 1407
QY 1381 CGTGTGTATCGGAAACAAAGTCAATGTTTGTATATPCTCATCAGGTCTCAGCTCGGTAT 1440
DB 1408 CGTGGCCATCGGAAACAAAGTCAAGTGCCTGTGTATCCCCATCAGGTCTCAGCTCGGTAT 1467
QY 1441 GATCAAGATACTGATATCTGAGGTGTGATTTTTCAGTCAATTAATTTGGGACATATTT 1500
DB 1468 GACCAAGATACCTGATATCTGAGGTGTGATTTTTCGCTCATATATTTGGGACATATTT 1527
QY 1501 TCTGGAGAAATGAACATATCTTCTGTGTTTCATGGTGTGAGATTACTCAACTTCTAGTT 1560
DB 1528 TCTGGAGAAATGAACATATCTTCTGTGTTTCATGGTGTGAGATCAACAACATTTCTGGTC 1587
QY 1561 CCACCTGAAACCTGTAGTGAAGATGACAGCTGCATCTGCTGTAGCCAGTGAACAC 1620
DB 1588 CCCCCGAAAACCTGTAGTGAAGATTTCAACCTGCATCTGTTCCGTAGCCAGTGAACAC 1647
QY 1621 TCAGTAGGACTTCTAATGTTTGCAGAGAAAATGATTAATGTTGGCATCTGCTCACCTT 1680
DB 1648 TCTGTAGGCTCTTAAGTCTGCGAGAGAAAATGCAATGTTTGGCGTCTGCGCACCTC 1707
QY 1681 TTTCCCTATTCAAGTAATCAATGAGGCTTCTGTGATTAATCACTGTGGTGGGCTGTCA 1740
DB 1708 TTTCCCTATTCAAGTGAATCAAGTGGAGGCTTCCGACGATCACTGTGGTGGGATGACCA 1767
QY 1741 GATGGTCTCTGTACGCTGTGGCAAAATGGATPACTGTGTGCAATGGATCGTTGTGTGAGGGG 1800
DB 1768 GACGGCTCTGTGATATGCTGGCAGATGACACACTGTGTGGCTGGATCGTTGTGCAATGGGA 1827
QY 1801 ATAAACGAGTGTGAGATTTCTAAACGCTTGTGATGAAGCTGTGCTGCTGCTGTGATTCA 1860
DB 1828 ATCACAGCAGTGGAGATCTGAAACGCTTGTGACGAAGCGGCTTCTCGCGCTGTAGACTCT 1887
QY 1861 CTTTGTATCCAGCAGTCAACCTTAAACAAAGCTATGACGAGACGTAGTCTTGTGCTCTT 1920

DB 1888 CTTTGTATCCAGCAGTCAACCTGAAACAAGCCATGACAAGACGAGTCTCGCTGCCCTT 1947
QY 1921 AAAAATATGGCCCATATAAGCTCAAAACCCCTTGCAACTTAACCTCTTGGCTTCTGAGCA 1980
DB 1948 AAAAATATGGCCCATATAAGCTCAAAACCCCTTGCAAGCAACCTTTGGCTTCTGAGGCC 2007
QY 1981 TCTCAAGAGGAAATTTTACCTAAATTTCTCATAACTCCCTGATGGTTCAGCAATAAAG 2040
DB 2008 TCTGAAAGGGTAATTTTACCTAAATTTCCCAATACTCCCTGATGGTTCAGCAATAAAG 2067
QY 2041 ACAAACTTAACAGACCCCGACATACATGTGCTATTCTTTGATGTGGAAGCGTTGATTT 2100
DB 2068 ACATAACCTTAACAGACCCCGATATCATGTGCTTTCTTTGATGTGGAAGCTTTGATTT 2127
QY 2101 CAACTCTCTGATGAAGAGCCCTTAGGGCGGATATCTGCTCTTATTTCGCCAGAGATTTG 2160
DB 2128 CAACTCTCTGATGAAGAGCCCTTAGGGCGGATATCTGCACTTATTTCGCCAGAGATCTG 2187
QY 2161 CAABAAGCATCTGGCAGTTTCAGACAAAGGGGCTCTTTTAACTGGAAAACGAGCAGCA 2220
DB 2188 CAGAAGCATCTGGCGATTCAGACAAAGGGGCTCTTTCTTAACCTGGGAAGCGAGCGG 2247
QY 2221 GTTCTCTTCCAACTAAGTGAAGAAACGATCAAGAGAAACATCAAGGAAACACCTCTGAT 2280
DB 2248 GTTCTCTTCCAGCAAGTGAAGAAACTATCAAGAGAAACATAAAGGAAACACCTCTGAT 2307
QY 2281 GATGAAGAGGAGGATGAGAGATTAATGAGGCGAGAGAGGAGGAGATAGTGACCTGATAT 2340
DB 2308 GAGGAGGAGGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2367
QY 2341 CGGTCCAGCAATCAAGCCATTGACCTTATTAGAATATAATTTAACTATGACACTGCA 2400
DB 2368 CGGGCAGCAAGTCCAGCCACTGACCTACTAGAAATACNACCTTACTATGATACCGCA 2427
QY 2401 AAGCTGTTTATGTCCTGCTTCAAGCTGGGGTTTGAATGAAGTACTGGATGAAGTTGC 2460
DB 2428 AATATTATCATGCTCTGCTCCAGCCTGGGGCTTGAATGAAGTCTGATGAAGTTGT 2487
QY 2461 CTGGATCGGCTTGGAAATGCTGAACCCCACTGACCGTATCGTTGGCTCTTGTCAAGA 2520
DB 2488 CTCGATCGCTCGGAATGCTGAACCCGCACTGACCGGTGCTCTTGGTCTCTATCGAGA 2547
QY 2521 GGAGGCCATATCTCATGATGCTGCGGGTTATAATCAGCCTGCTGTGTAACCTGTCACAT 2580
DB 2548 GGAGTCAATGTCCTGATGCTTCCGGTTATATCAGGCTGCTGGAACCTGCTGCTGAT 2607
QY 2581 GGGAAAACAGAAAGTAGGAAGCTGCCAGCTGTGAGGGAGTAGGAAGGGAACCTTAC 2640
DB 2608 GCGAAAGCAGAAAGTAGGAAGGAGCTGCCAGCAGCGGAGGCGTAGGCAAGGGAACCTAC 2667
QY 2641 GGAGTGTCCGTCGCTGACCAACAGCATCTCTGCTGCTATCATTTCTTTGGCAATACT 2700
DB 2668 ACAGTGTCTCGAGCGGTCACTCAGCATCTGTTGTCATCATATCCCTGGCGAATACT 2727
QY 2701 TTAATGAGTATGACCAATGCAACTTTTATTGTGTGATCATATGAAGAGGGTCTACACAG 2760
DB 2728 TTAATGAGTATGACCAATGCAACTTTTATTGTGTGATCATATGAAGAGGGGCCACAG 2787
QY 2761 CCACCTTAGAACCAAGCACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAACTTCCAGT 2820
DB 2788 CCGCTTAGACCTGGCACCCAGACCTCTCTAAGGGGAGGGAATTCCTCCAGCCTCCAGT 2847
QY 2821 AATATTGTGCAAGGACAGATTAACAAGTGTGTCACCTGCTGTTTCCGCTCGGCTGAT 2880
DB 2848 AACATTGTGCAAGGACAGATTAACAAGCCGCTGGCGCTGTGCTGCTGCTGCTGAT 2907
QY 2881 GCTGATCACTCGGCTCTGACCCCTCTCTGCTCTGCTCTTACATACCTGTTTCTTAGTA 2940
DB 2908 GCTGATCACTCGGCTCTGACCTCTGCTCTGCTCTGCTCTTACCTACCTGTTTCTTAGTA 2964
QY 2941 AATGAAGGTTGGAGTCAGTTAGTGTGATGCACTGTGTTATTGCTGCGACACCTTCTGGGA 3000


```
Db 2965 AATGAAGCTGGAGCCAGTTAGCGCCATGCACTGTGTTATGTTGCTGACCTGCTGGGG 3024
Qy 3001 TTGGATAAATTTAGGCTCCCTTCTGGAGATGCTGCGCCGAGATGCGAAGATGCAATGC 3060
Db 3025 CTGGAGAGATTAGGCTCTCTCTCTGGAGATGCTAGCTCGAAGATGCGCAGGACCGATGC 3084
Qy 3061 TTGGAGGTGAGAGAGCCGACAGAGCCCTCTCTCTGGCGAACTGGAAGAAATTGAGCAG 3120
Db 3085 TTGGAGGTGAGAGAGCTGACAGGCCCTGCTCTTAGCTGAGCTGAGNAGGATTGAGCAG 3144
Qy 3121 GCAGGCGAGGAAGGCCATTGATGCTGGGCTCTTACTTACTCAGTACATAGACCAC 3180
Db 3145 GCAGGCGAGNAGGAGACGATGACACTGGGCTCTTACTTACTCAGTATAGACCAT 3204
Qy 3181 GTCATATCACTGGAGTCAATCAGAAAGCGGCGACATATCACACCGCTCTCTGATGCC 3240
Db 3205 GTCATATCACTGGGCTCACGGCGAAGCCATGACAGCTATGCGAGCTGCTCCAGATGCC 3264
Qy 3241 TCAGGCGCTGAAGCAAAAGTCCAGGAGGAAGAGCATGACTTGTGACGATGACATCAC 3300
Db 3265 TCAGGCGCCAGAACCCAAAGTCCAGGAGGAAGAGCATGACTTGTGATGATGATCATCACT 3324
Qy 3301 ACTGGTTGCTTATCAAGTGTCCACAAATGAAAAAATTTCTACATCTTACGAGGAAGA 3360
Db 3325 GCAGGTGCTTGTCAAGTGTCCACAGATGAAGAAATCTCCAGCTTTCAGAGAAAGA 3384
Qy 3361 CGGAAGCAAGTACCGCTATTGTTTACTTGGAGTAATAGGAGCTGAATTTGGTGCTGAA 3420
Db 3385 AGGAGCAGGCCACGCTATTGTTCTCTGGAGTGATAGGGCGAGGTTTGGAGCTGAA 3444
Qy 3421 ATTGAACCTCTTAACCTATTGACAGACCTCGAAGCTTAGCCAAATTTCTGAGGATTC 3480
Db 3445 ATTGAACCCGCAAGCTGTGACAGGCTCGAAGCTCGAAGTCAATTCCTGAAGCTTT 3504
Qy 3481 GGGTTGACTAGTGGTGGATCCCACTACTCGTGGCCAGACATCTTTCGAAAGCACTGACG 3540
Db 3505 GGGTTGACCGAGGAGGCTCCCACTACTCCCTGGCCAGACACACGTGCAAGGCACTGACA 3564
Qy 3541 TTTCTTCTCTACAGCTCCAAAGCCCAAACTTCTCCACACAGACATCTCCGAAGACA 3600
Db 3565 TATCTCTGTACAGCCACCAAGTCCCAAGTCCCTCTCTCATAGCACCATCCGGAGACT 3624
Qy 3601 GCATGATCTGANTGGAAGTGCGGTTCACTGTTTGGAGGCTTTACATGATGTGTCGGCT 3660
Db 3625 GCCACTGACTGATTGGGCGAGGTTCACTGTATGGAGGCTTACATGATGTGTCGGCT 3684
Qy 3661 GTTCTGATGGGCTTCTGAACTTTGTGCGGATGCGGAAACAACTTGCACACATCACA 3720
Db 3685 GTTCTGATGGGCTGCTGGAGCTCTGTGAGAGCTGTGAGAGAGCTGAGAAACAACTTGCACACATCACA 3744
Qy 3721 ATGGGGTTGCTCTGAGCCAGCAGCTGACTCGGCCGCTCTGCGAGGCAATGCCCTCTCG 3780
Db 3745 ATGGGGTGTCTTGTAGCCCTCGAGCTGACTGTGCGGCTCGCAGGCAAGCCCTGTCT 3804
Qy 3781 CTCTATGTCACCGCCAGACCAACCGCTTTCATCACCAACATAGCCAAAGAGGTACACAGA 3840
Db 3805 CTCTATGTCACCGCCAGACCAACCGCTTTCATCACCAACATAGTAAAGGAGGTCCACAGA 3864
Qy 3841 CATACGGCTCTTGCAGCAAAATACCCATCACAGCAGAAATATGCACACAACTCTTGA 3900
Db 3865 CACACGGGCTTTCGACGCAACACCCAGTCCAGCAGAGATATACACACCAACCCCTGGCG 3924
Qy 3901 CGAGCTAAAGGGGAAATTTGAGAGTCAATTGAAATCTTATTGAAAGATGCCACAGAT 3960
Db 3925 AGGGCTAAAGGGGAAATCTGAGAGTCAATTGAAATCTTATTGAAAGATGCCACGAT 3984
Qy 3961 GTTGTGGATCTTCTGTGAGGATTAAGACATCATTAATGATCTGCTTGAAGGATCTTTA 4020
Db 3985 GTGTGGATCTTCTGTGAGGATTAAGATATCATCATGTACTGCTGCTGGAAGGATCTTTA 4044
Qy 4021 GTTAAAGAAAGGCTTCAAGAAATGTTTCCAGCCATCTGCAAGGTTTACATGTCAGC 4080
Db 4045 GTGAAAAAGAGGCTTTCAGGAGTGTGTTTCCAGCTATCTGCAAGGTTTCTACATGTCAGC 4104
```

```
Qy 4081 TATTATGAGCGGAATACAGAAATAGCAGTTGGAGCTGCGCATGCTTCACTGGCCCTGTATC 4140
Db 4105 TATTATGAGCGCAGTACAGAAATCGCAGTTGGAGCAGCGCATGGCTCAGTAGCCCTGTAT 4164
Qy 4141 GACATCCGAGCTGGAAATGTGAGAAATCCATGACACAAAGGACCAATCACTGAGTG 4200
Db 4165 GACATCCGAGCTGGAAATGTGAGAAATCCATGAGGACCAATCACTGAGGCGTG 4224
Qy 4201 GCTTTTGCTCTGATGGAATATCTTGCCACTACTCAAAACATGACACCAATTTCT 4260
Db 4225 TCCTTTGCTCTGATGAGCGGTTTACCTTGCCACTACTCGAACACCGACACCAATTTCC 4284
Qy 4261 TTTTGGCAGATGAACACGTCACCTGCTGCGAAGCATGCGCATGCTGAACCTGCGACCTCAG 4320
Db 4285 TTTTGGCAGATGAACACCTCACTCTCTGGGAAGCATTTGAATGCTGAACCTCAGCTCCTCAG 4344
Qy 4321 CTGCGCTGCATTAACCTTACCAGTGCCCTGTCGACCGCGCTGCCCCGGCTCCCCAC 4380
Db 4345 CTGCGCTGCATTAACCTTACCAGTGCCCTTCCAGTGACCGCAGCATCCCCGGCTCCCCAC 4404
Qy 4381 AATGCCCTCAAGCTGCGCGGCTCATCTGGACTTCCAAACCGCAAGCTCATCTCATGGCC 4440
Db 4405 AACGCCCTTAAGTTGCGCGGCTCATCTGGACTTCCAAACCGCAAGTATCTCTCATGGCC 4464
Qy 4441 CATGACGGGAAGGACACCGCTTTCATGCTTAA 4473
Db 4465 CATGATGGGAAGAACACCGCTTTCATGCTTGA 4497
```

RESULT 5
AF188124

LOCUS Mus musculus TGF-beta resistance-associated protein TRAG (Trag) 467 bp mRNA linear ROD 15-SEP-2001

DEFINITION mRNA, partial cds.

ACCESSION AF188124

VERSION AF188124.1 GI:15624072

KEYWORDS Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 4467)

AUTHORS Sanders, S. and Thorgeirsson, S.S.

TITLE TRAG: a novel gene associated with TGF-beta resistance

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 4467)

AUTHORS Sanders, S. and Thorgeirsson, S.S.

TITLE Direct Submission

JOURNAL Submitted (20-SEP-1999) LEC, NIH, 37 Convent Drive, Bldg. 37, Rm

3C28, Bethesda, MD 20892-4255, USA

FEATURES

Location/Qualifiers

1..4467

/organism="Mus musculus"

/mol_type="mRNA"

/strain="129SvJ"

/db_xref="taxon:10090"

/chromosome="18"

/map="18D.1-E.3"

/tissue_type="brain"

1..>4467

/gene="Trag"

1..>4467

/codon_start=1

/product="TGF-beta resistance-associated protein TRAG"

/protein_id="AAL03982.1"

/db_xref="GI:15624073"

/translation="MAGNSLVLPVLMGRKAPTHCISILLTDDGGTIVTGDGIC

LWDVSELEVNPRALLFGHTASITCLKACAGDKRYTVTSASANGEMCLWDVDCRCI

EFTKLACTHTGIGTFQFVSGNQGRLLCHGYPEILVVDATSLVLSVSKISPDW

ISSMSIHSQRTQEDTVVSLGILKVMIVTSEMSGMDTEPIFEESKPIYCNQC

SISFCAPTQSRLLVVCYSKYWRVFDAGDYSLLCSGFSNGQWTGTGDFVSNADKVIITE

NGOSVIYKLPASCLPSPDSFRSDVGKAVENLIPVQHSLLDQKDKELVICPVTRFFY
GCKEYHLKLLTQDSSGRNLNWNADIAEKQADBLXWTTCTISLQEAFLDKLKPAG
I1DQUSV1PNSNEPKVTASVYIPAHGRVLCRGEDGSI11VPATQI1VLQ1QGEHML
RGMWPHRTLRHNRKNTCLLYPHQVSARYDQRYLISGGVDFSVI1WD1FSGEMKHF
CVHGE1TQLV1PPENCVRQH0C1CSVASDHSVGLLSREKKKCI1MLASRHLFP1QVI
KWRPDD1YLVGS2DGSVVWOMD7GALDR0CAMG1TAVE1ILNACDEAVPA1VDSLSHP
AVNLKQAMTRSLAALKNMAHHLK1OTLATNLLASBASDKGNL1PKYSHNSLMVQ1KTN
L7DPDIH1FVFEAL11I1QLLT1EASRPN1TALL1SPENT1QASGSDSKGGS1FLTKRAA
VLFOQVKT1KEN11KEHLLDEED1BEEAR1R0RSD1DEY1RASRSK1PULLE1N1TMD
TAK1FMSCL1HAWGLNE1DEV1CLDR1LMLKPH1CTVSPGL1SRGHMS1LMLPY1NOAG
KLLHAK1EAG1RKL1PAE1GVKG1TVSRA1VTTQH1LLS1I1SLANT1MTNAT1F1GDHM
KKGP1RPR1GPT1DLSK1ARDS1PASS1NVQ1Q1K0AA1PVVS1ARD1DHS1GSD1SPA
L1PCT1FVNEGS1Q1AM1C1VLP1DL1GLER1PR1P1L1EMLAR1WDR1CLEV1RE1QA1LL
L1BLR11SOGR1K1ET1DW1P1YLP1Q1MDH1V1SPG1TAE1AM1Q1TAA1P1AS1PEAK1VQ
E1EHL1VDD1D1TAG1CLSS1VP0MK1IST1SEYER1R0KAT1A1VL1GV1GAE1F1E1P1PKLL
TRP1RSS1Q1PE1G1T1SGS1NS1LAR1HT1CKAL1TYLL1Q1P1P1K1P1P1P1T1R1T1AD1LI
GRG1FTW1P1YMD1SAV1MGL1EL1CAD1B1K1Q1LAN1T1MGL1PL1PA1D1S1ARS1AR1H1SL1IA
TAR1P1AF1T1I1AK1E1VR1HT1A1LA1NT1Q1S0Q1I1HT1TLAR1AK1E1IL1R1V1B1I1EK1MPT1DV
VDLL1VE1MD1I1MY1CLE1G1SLV1KK1G1LQ1E1CP1AL1CR1F1V1Y1VS1R1H1R1I1AV1GAR1H1G1S1VAL
YD1RT1G1CK1Q1T1I1GH1G1P1T1AV1SP1AF1D1GR1Y1LAT1YS1NT1D1SH1I1SF1W1Q1M1NT1S1IL1GS1GM1NS
APQLRC1KI1TY1Q1VP1Q1P1AS1P1G1PS1H1N1ALK1LAR1LI1W1TS1NR1NV1IL1WA1HD1G1K1E1H1RF1W1"

ORIGIN

Query Match		79.3%;	Score 3546.8;	DB 10;	Length 4467;
Best Local Similarity		87.2%;	Pred. No. 0;		
Matches 3900;		Conservative 0;	Mismatches 567;	Indels 3;	Gaps 1;
Qy	1	ATGGCAGGAAACAGCGCTTGTCTACCCATTTGTTTGGGGTCGAAAGCGGCCACACAT	60		
Db	1	ATGGCAGGAAACAGCGTAGTCTTGCCCATTTGTTTGGGGCGCAAGCACCCACACAT	60		
Qy	61	TGATCTCAGCGGTACTTTTAACAGATGATGGGGCCACGATCGTAACAGGATGTCAAGC	120		
Db	61	TGCAATTCGTCAATACTGTGTGAAGATGATGGGGCCACAAATTTGTAATCGATGSCACGAT	120		
Qy	121	GGACAAATATGTCTCGGGATCTTTTCAGTAGAACTTGCAAAATTAATCTCGAGCACTGTGTG	180		
Db	121	GGACAAATATGTCTCGGGATGTTTCGGTAGAACTTAGAAGTTAATCCCGAGCACTGTTA	180		
Qy	181	TTTGGTCATACAGCATCAATCACTGTTTGTCTAAAGCTTGTGCTTCCAGTGCAACAAACAG	240		
Db	181	TTTGGCCACACAGCATCCATCACTGTTTGTCAAAAGCCTGCGCTTCTGGGGACAAAGCG	240		
Qy	241	TATATTGTGAGTGATCTGAAAGTGGAGAGATGTGCTCTGGATGTGAGTGGCAGCA	300		
Db	241	TACACTGTGAGCGGCTCTCAAAACGGAGAGATGTGCTCTGGATGTGAAACGATGGCAGA	300		
Qy	301	TGTATTGAATTTACAAAATTTAGCTGCACACATCTAGGCATACAGTTCTACCAAGTTCTCT	360		
Db	301	TGTAATTGAATTTACCAAGTTAGCTGCACACACATCGGCATATACAGTTCTACCAAGTTCTCT	360		
Qy	361	GTGGGAATCAGCGAAGGAGGCTTTTATGCCAGGACATTAACCTCGAAATTCCTTGT	420		
Db	361	GTGGGAAACGACAGAGGCGAGCTCCTCTGCCATGGACATTAACCTCGAAATTCCTCGTT	420		
Qy	421	GTGGATGCTACAGCCTTGAAGTATTAATCTCTCTAGTATCAAGATATCAACAGACTGG	480		
Db	421	GTGGATGCCACCAGCCTTGAAGTGTGTATTCCTTGGTATCGAAGATCTCTCAGACTGG	480		
Qy	481	ATTAGCTCATGATGATTAATTCGATCCCAAGCAACAAAGAGGACAGTGGTAGCACTC	540		
Db	481	ATCAGCTCATGATGATCATCCACTCTCAGCGGACAAAGAGGACACTGTGGTGGCGCTC	540		
Qy	541	TCGGTGAATCCTCGAGGCTGAGGTTGTGTTTACCTCGGAAATTAAGTACATGCAAGAT	600		
Db	541	TCGTGACAGGTTATCTGAAGGTGTGGAATTTGTCAACCTCTGAAATGAGTGAATGCAAGAT	600		
Qy	601	ACTGAGCCAAATTTTGAAGGAGGATCCAAACCAATTTATGTGAGATTTGCCAAGCATC	660		
Db	601	ACTGAGCCAAATTTTGAAGGAGGATCCAAACCAATTTATTTGTGAGAAATTTGCCAAGCATC	660		
Qy	661	TCCTTTTGTGATTTACAAAGGTCACATTTTGGTGTGTGTTTCCAAATTAATGGAGGTTG	720		

Db	661	TCCTTTTGTGATTTACACAGAGGTCACATTTTGGTGTGTGTTCCAAATTAATGGAGGTTG	720
Qy	721	TTCCGATCCGAGACTATTCCTCTGTTGTTTCAGGTCCTAGTGAAAATGGACAGCATGG	780
Db	721	TTTGATGTCGCGACTACTCTCTGTTGTGCTCAGGTCCTAGTGAAAATGGACAGCATGG	780
Qy	781	ACCGGGGGGACTTTGTCTCATCAGATAAAGTCATCATTTTGGACAGAAAATGGGCAAAAGT	840
Db	781	ACTGGAGGGGACTTTGTGCTGCGAGACAAAGTCATCATCTGGACAGAAAATGGGCAAGAT	840
Qy	841	TATATTTACAAACTACCTGCCAGTTGCTTCCAGCTAGTGAATCAATCCGCAAGTATGTTG	900
Db	841	TACATCTACAAACTCCCTGCCAGTTGCTTCCAGCTAGTGAATCAATCCGCAAGTATGTTG	900
Qy	901	GGGAAGGAGTGAATAATTAATTCCTCTCTGTAACAATATCTCTTGGATCGAAAAGAT	960
Db	901	GGGAAGGAGTGGAAAATTTGATCCCTCTCTGTCAGCATAGCCCTCTTGGATCGAAAAGAT	960
Qy	961	AAAGAGTTGTAAATTTGTCTCTGTTACTCGGTTCTCTATGATGATGACAGAAATATTTTC	1020
Db	961	AAAGAGTTGTGAATTTGTCTCTGTTACTCGGTTCTCTATGATGATGACAGAAATATTTTC	1020
Qy	1021	CATAAATCTGTTAAATTCAGGGTGAATTTCTCTGGAAGTTGAATATTTGGAACATATCAGAC	1080
Db	1021	CATRAAGCTACTAAATTCAGGGCGATTTCTCTGGAAGTTAAATATCTGGAACATATCAGAC	1080
Qy	1081	ACAGCTGATAACAGGGAAGTGAAGNAGGCTGCGCAATGACAACTCTCTATTAGTTTGCAA	1140
Db	1081	ATAGCAGAGAACAAGGAAGCCGATGAAGGGCTAAAGATGAACAATTTGATTAGCTTGCAA	1140
Qy	1141	GAGGCATTTGATAAATCTGAATCTTGTCTCTGGAATTTATAGATCACTCAGTGTGATTT	1200
Db	1141	GAGGCATTTGACAGCTGAGCCCTTGCCTGCTGGAATTTATCGATCAGTGTGATTT	1200
Qy	1201	CCCATAAGTAAGAACCTCTTAAAGTAACTGCAAGTGTGTACATACCAGCACATGGAAGA	1260
Db	1201	CCCAACGAATGAACCTCTTAAAGTAACTGCGAGTGTCTACATACCAGCACACGCGGGA	1260
Qy	1261	CTTGTGTTGCTGCTGGAAGTGGAGCATAGTTATTTGCTTACCTGCCACACAGACGCGCAT	1320
Db	1261	CTTGTGTTGCTGCGGGAAGTGGAGCATCATTTATTTGCTGCTGCCACACAGACGCGCAT	1320
Qy	1321	GTACAGCTGTTGCAAGGGGAAACACATGCTCAGAGAGGTTGGCCACCTCACAGAAACATCT	1380
Db	1321	GTTCAGCTATTGCAAGAGGAAACACATGCTCAGAGAGGTTGGCCCTCACAGAAACCTCT	1380
Qy	1381	CGTGTCATCGGAAACAAAGTCAATGTTTGTCTATTCCTCATCAGGTCCTCAGTCTCGGTAT	1440
Db	1381	CGTGCCCATCGGAAACAAAGTCACTGCTGTATTCCTCATCAGGTCCTCAGTCTCGGTAT	1440
Qy	1441	GATCAAGATACCTGATATCTGGAGGTGGAATTTTTCAGTCATATTTGGGACATATTT	1500
Db	1441	GACCAAGATACCTGATATCTGGAGGTGGAATTTTCCGTCATATTTGGGACATATTT	1500
Qy	1501	TCGTCAGAAAATGAAAACATATCTTCTGTGTTTATGTTGTTGAGATTTACTCAACTCTAGTT	1560
Db	1501	TCGTCAGAAAATGAAAACATATCTTCTGTGTTTATGTTGTTGAGATTCACACACTCTCTGTC	1560
Qy	1561	CCACCTGAAAATCTGTAGTCAAGATAAGCACTGCTGCTCTGTAGGCAAGTACACAC	1620
Db	1561	CCCCCGGAAAACCTGTAGTCAAGATTTCAACACTGCTGTTTCCGTAGCCAGTACACAC	1620
Qy	1621	TCAGTAGGACTTCTAAGTTTTCGAGAGAAAATGCAATGTTGGCATCTCGTCACTTT	1680
Db	1621	TCGTAGGCTCTTAAAGTCTGCGAGAGAAAATGCAATGTTGGCGTCTCGCCACCTCT	1680
Qy	1681	TTTCTCTATTCAAGTAATCAAAATGGAGGCTCTCTGATGATTAACCTGTTGGTGGGTTTCA	1740
Db	1681	TTTCTCTATTCAAGTAATCAAAATGGAGGCTCTCGAGACTACTCTGTTGGTGGGATCCACA	1740
Qy	1741	GATGTTCTGTGTACGTCGTCGCAAAATGGATACTGTTGATGATTCGTTGTGTGATGGG	1800
Db	1741	GACGCTCTGTGTATGTCTGCGAGATGGACACTGTTGGGCTGGATCGTTGTGCAATGGGA	1800

Qy	1801	ATAACAGCAGTTGAGATTTCTAAACGGCTTGTGAAGAGCTGTTCCCTGCCTGCTGTTGATTCA	1860
Dd	1801	ATCACAGCAGTGAGAGATCTGAACCGCTTGTGAACGAGCGTTTCCTGCCGCTGTAGACTCT	1860
Qy	1861	CTTAGTCAATCCAGCAGTCAACCTTAACAAGCTATGACGAGACGTAGTCTTGTGCTGCTTT	1920
Dd	1861	CTTAGTCAATCCAGCAGTCAACCTTAACAAGCTATGACGAGACGTAGTCTTGTGCTGCTTT	1920
Qy	1921	AAAAATATGCCCATTATAGCTACAAAACCTTTGCAAATTAACCTCTTGGCTTCTGAGGCA	1980
Dd	1921	AAAAATATGCCCATTATAGCTACAAAACCTTTGCAAATTAACCTCTTGGCTTCTGAGGCC	1980
Qy	1981	TCTGACAGGGAAATTTACTAATAATTTCTATAATCTCTCATATCTCCCTGATGGTTCAAGCAATAAAG	2040
Dd	1981	TCTGACAGGGAAATTTACTAATAATTTCTATAATCTCTCATATCTCCCTGATGGTTCAAGCAATAAAG	2040
Qy	2041	ACAACTAACACACCGGACATACATGCTATTCTTTGATGTTGGAAGCGTTCATATTAT	2100
Dd	2041	ACAACTAACACACCGGATATCCATGCTTTTTCTTTGATGTTGGAAGCTTTGATATT	2100
Qy	2101	CAACTCTGACTGAAGAAGCTCTAGGCCGAATACTGCTCTTATTTCCTCCAGAGAAATTTG	2160
Dd	2101	CAACTCTGACTGAAGAAGCTCTAGGCCGAATACTGCACTTATTTCCTCCAGAGANTCTG	2160
Qy	2161	CAAAAAGCATCTGGCAGTTTCAGACAAAGGGGCTCTTTTTTAACTGGAAAAACGACGACCA	2220
Dd	2161	CAGAAAGCATCTGGCAGTTTCAGACAAAGGGGCTCTTTTCTTAACTGGGAAGCAGCAGOG	2220
Qy	2221	GTTTCTCTCCMAAGTGAAGAACGATCAAAAGACATCAAGGAACACTCTCTTGAT	2280
Dd	2221	GTTTCTTTTCCAGCMAGTGAAGAACTATCAAAAGAGACATAAAGGAAACACTCTCTGGAT	2280
Qy	2281	GATGAAGAGAGATCAGAGATATATGAGCAGAGAGAGGAAGAAAGTATCTCTGAATAT	2340
Dd	2281	GAGAGAGAGA CAGAGAGAGGGAGAGAGCAGAGCAGGAGGATATGATGACCTTGAGTAC	2340
Qy	2341	CGTCCAGCAAACTCAAGCCATTGACCCTTATTAGAAATATAATTTAACTATGGAACACTGCA	2400
Dd	2341	CGGGCCAGCAAGTCCAAGCCACTGACCCTTACTAGAAATACAACCTTACTATGATACCGCA	2400
Qy	2401	AAGCTGTTATGCTTCCTTCAGCGCTGGGTTTGAATGAAGTACTGGATGAAGTTTG	2460
Dd	2401	AAATTTATCATGTCCTCTCCACGCTGGGCTTTGAATGAAGTTCTGATGAAGTTTGT	2460
Qy	2461	CTGATGCTGCTTGGATGCTGAAACCCCACTGACCGTATCGTTTGGCTCTCTGCTCAAGA	2520
Dd	2461	CTGATGCTGCTTGGATGCTGAAACCGCACTGACCGTATCGTTTGGCTCTCTTATCGAGA	2520
Qy	2521	GGAGGCCATATGCTCACTGATGCTGGGGTTATATATCAGCCTGCTTGTAAACTGTCACAT	2580
Dd	2521	GGAGGTCATATGCTGATGCTTCCGGTTATATCAGGCTGCTGGGAACCTGCTGCAT	2580
Qy	2581	GGGAAACAGAGTAGGAAGAGAGCTGCCAGCGTCTGAGGGAGTAGGAAGAGGAACCTTAC	2640
Dd	2581	GCFAAACAGAGTAGGAAGAGAGCTGCCAGCAGCGGAGGGCGTAGGCAAGGGGAACCTAC	2640
Qy	2641	GGAGTCTCCGCTGCTGACCAACAGCATCTCTGCTATCATTTCTTTGGCAAAATACT	2700
Dd	2641	ACAGTGTCTGAGCGGTACCACTCAGCATCTGTTGTCATCATATCTCTGGCGAATACT	2700
Qy	2701	TTAATGAGTATGACCAATGCAACTTTTATTTGTTGATCATATGAAGAGGGTCTTACCAGG	2760
Dd	2701	TTAATGAGTATGACCAATGCAACTTTTATTTGCGATCATATGAAGAGGGGCCCAACAGG	2760
Qy	2761	CCACTTAGACCAAGCACCCAGACTTTCTTAAGGCAAGGGGTTCCCTCCAACTTCCAGT	2820
Dd	2761	CCGCTTAGACTGGCACCCAGACTCTCTAAGGCGAGGGATTTCCCTCCAGCCTCCAGT	2820
Qy	2821	AATATTGTCMAGGACAGATTAACAAAGTTGCTGCACTGCTGCTTCCGCTCGCTCTGAT	2880
Dd	2821	AACATTTGTCMAGGACAGATTAACAAAGCGCTGCTGCTGCTTCTGCTCGCTCTGAC	2880

Qy 601 ACTGAGCCAAATTTTGGAGGAATCCAAACCAATTTATTTGTCAGAAATGCCAAGCATC 660
Db 601 ACTGAGCCAAATTTTGGAGGAATCCAAACCAATTTATTTGTCAGAAATGCCAAGCCTC 660
Qy 661 TCTTTTGTGCAATTTACACAAAGCTCACTTTTGTGTGTGTTCCAAATATTTGAGGGTG 720
Db 661 TCTTTTGTGCAATTTACACAGAGTCTGCTTTGTGTGTGTTCCAAATATTTGAGGGTG 720
Qy 721 TTGATGCGGAGACTATTTCTTTGTGTGTGTTGAGTCTAGTCAAGATGACAGACATGG 780
Db 721 TTGATGCTGGCACTACTCTCTCTGTGTTTCCAGTCTAGTCAAGATGACAGACATGG 780
Qy 781 ACCGGGGGAGCTTTTGTCTCATCAGATTAAGTCAATTTGGACAGAAATGGCAAAAT 840
Db 781 ACTGGAGGGAGCTTTTGTGTCTGACAGAAAGTCAATTTTGGACTGAAACCGGGCAGAT 840
Qy 841 TATATTTACAAACTTACCTGCGAGTTGCTTCCAGTCTAGTCAATTTCCGCAAGTGTG 900
Db 841 TACATTTACAAACTTCCCTGCGAGTTGCTTCCAGTCTAGTCAATTTCCGCAAGTGTG 900
Qy 901 GGGAGGCGAGTTGAAATTTAATTTCTCTGTGTACAAATATCTCTTTGGATCGAAAGAT 960
Db 901 GGGAAAGCAGTTGAAATCTGATTTCTCTCCGTGCGAGCATAGCCTCTTTGGATCAGAGAT 960
Qy 961 AAGAGTTGCTAATTTCT 1020
Db 961 AGAGAGTTGTAATTTCT 1020
Qy 1021 CATAAACTGTTAATTTCAAGGTGATTTCTCTGGAAGTTGTAATTTTGGACATATCAGAC 1080
Db 1021 CATAACTACTAATTTCAAGGTGATTTCTCTGGAAGTTGTAATTTTGGACATATCAGAC 1080
Qy 1081 ACAGCTGATAAACAGGAAGTGAAGAGGCTGSCAATGACAACTTCTATTAGTTGSCAA 1140
Db 1081 ATAGCAGACAAACAGGAAGCAATGAAGGCTGAAAGAGCACTTCTGTTATTAGTTGSCAA 1140
Qy 1141 GAGGCAATTTGATTAACCTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
Db 1141 GATGCAATTTGACAACTGAAAGCCTTGTCTCTGGAATTTATCGATCAGCTGAGTGTGAT 1200
Qy 1201 CCCAATAGTAATGAACCTTTAAAGTAACTGCAAGTGTGTACATACAGACATGACGA 1260
Db 1201 CCCAAGAGCAACGAACTTTAAAGTAACTGCAAGTGTGTACATACAGACATGACGCGC 1260
Qy 1261 CTTGTTTGTGTCGTGAAGTGAAGCATAGTTATTTGATCTGCGCAACAGAGGCGCAT 1320
Db 1261 CTTGTTTGTGCGCGGGAAGACGAGCATCATTTATGTCCTCTGCGCAACAGAGGCGCAT 1320
Qy 1321 GTACAGCTGTTGAGGGGAAACATCTCTCAGAAAGGTTGGCCACTCAGACAACTC 1380
Db 1321 GTTCAGCTGTGCGAGGAGAACATCTCTCAGACGAGGTTGGCGCCCGCAGAACCCCTC 1380
Qy 1381 CGTGTCTATCGGAACAAAGTCAATGTTTCTATATCTCATCAGGTCTCAGCTCGGTAT 1440
Db 1381 CGTGGCCACCGGAACAAAGTCAATGTTTCTGTATCTCTCATCAGGTCTCAGCTCGGTAT 1440
Qy 1441 GATCAAGATACCTGATATCTGAGGTGTGGATTTTTCAGTCAATATTTGGGACATATTT 1500
Db 1441 GATCAAGATACCTGATATCTGAGGTGTGGATTTTTCAGTCAATATTTGGGACATATTT 1500
Qy 1501 TCTGGAGAAATGAACATATCTTCTGTGTGTGTTCAATGTTGGTGTGAGATTAATCACTTCTAGTT 1560
Db 1501 TCCGGAGAAATGAACATATCTTCTGTGTGTGTTCAATGTTGGTGTGAGATTAATCACTTCTAGTT 1560
Qy 1561 CCACCTGAAATCTGAGTGAAGTACAGCATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
Db 1561 CCGCAGAAATCTGATGCAAGAGTTTCAACATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
Qy 1621 TCAGTAGGACTTCTAAGTTTTCGAGAGAAATATGCAATATTTGGCATCTCTCTCTCTCTCTCT 1680
Db 1621 TCTGTAGGCTGTCTAAGTTTTCGAGAGAAATATGCAATATTTGGCATCTCTCTCTCTCTCTCT 1680
Qy 1681 TTTCTTATTCAAGTAATCAAAATGGAGGCTTCTCTGATGATTTACCTGTGTGTGGGTGTCTCA 1740

Db 1681 TTCCCTATTTCAGGTGATCAAGTGGAGGCTTCTGACGACTTACCTGTGTGTGGGTGTCACG 1740
Qy 1741 GATGGTTCTGTGTACGCTCTGGCAAAATGGAATCTGGTGCATTTGGATCGTTGTGTGATGGG 1800
Db 1741 GACGGCTCTGTGTGTGTGTGCGCAGATGGACACTGGTGGCGCTGACCGCTGTGCAATGGG 1800
Qy 1801 ATAAACAGCAGTGTGAGATTTCTAAACGCTGTGTGTGTAAGAGCTGTTCTGTCTGTGTGATCA 1860
Db 1801 ATAAACAGCAGTGTGAGATTTCTCAATGCTGTGTGACGAAGCTGCTCCCTGTCAGCAGTGAATCA 1860
Qy 1861 CTTTGTGTCATCCAGCAGTCAACCTTAAACAGCTATGACGAGAGCTAGTCTTGTCTGTCTCTT 1920
Db 1861 CTGAGTCACCCAGCAGTCAACCTTGAAGCAAGCCTATGACGCGGAGTCTCGCGCGCCCTT 1920
Qy 1921 AAAAATATGGCCCATCATTAAGCTTACAAACCTTTGCAACTTAACTCTTTGGCTTCTGAGGCA 1980
Db 1921 AAAACATGSCCCACCAACAGCTGCAAACTTCTGCAACTTAACTCTTTGGCTTCTGAGGCC 1980
Qy 1981 TCTGACAGGGAAATTTTACCTAAATATTTCTCAAACTCTCCCTGATGGTTCAAGCAATAAG 2040
Db 1981 TCTGACAGGGAAATTTTACCTAAATATTTCTCAAACTCTCCCTGATGGTTCAAGCAATAAG 2040
Qy 2041 ACAACCTAACACAGCCGAGACATACATGTCTATTTCTTGTATGTCGGAAGCTTGTATTT 2100
Db 2041 ACAACCTAACCTGACCCGAGATATCCATGTCTTTCTTGTATGTCGGAAGCTTGTATTT 2100
Qy 2101 CAATCTCTGACTGAAGAGCCTCTAGGCCGAATACCTGCTCTTATTTCTCCCGCAGAGAAATTTG 2160
Db 2101 CAATCTCTGACTGAAGAGCCTCTAGGCCGAATACCTGCTCTTATTTCTCCCGCAGAGAACTG 2160
Qy 2161 CAAAAAGCATCTGGCAGTTTCAGAAAGGGGCTCTTTTTTAACTGGAAGAACGAGCAGCA 2220
Db 2161 CAGAAAGCATCTGGCAGTTTCAGAAAGGGGCTCTTTCTGACTGCAAGAACGAGCAGCA 2220
Qy 2221 GTTCTCTTCCAACTGAAGAAAGCAATCAAGAGAAACATCAAGNACACCTCTCTTGTAT 2280
Db 2221 GTTCTCTTCCAACTGAAGAAAGCAATCAAGAGAAACATCAAGNACACCTCTCTTGTAT 2280
Qy 2281 GATGAAGAGCAGGATCAGGAGATAATGAGGACAGAGAGGAGGAGAAAGTGTATCTGGAATAT 2340
Db 2281 GAGGAGAGGAGCAGGAGAGAGGATGAGGACAGAGGAGGAGAGAGAGTGTATCTGAGTAC 2340
Qy 2341 CGGTCCAGCAAACTCAAAGCATTGACCCCTATTAGAATATAATTTAACTATGGAACACTGCA 2400
Db 2341 CGGGCCAGCAAGTCCAAACCACTTACCTACTAGATACAACTTACTATGATACCGCA 2400
Qy 2401 AAGCTGTTTATGTCCTGCTTCAAGCCTGGGGTTTGAATGAAGTACTGGAATGAAGTTTGC 2460
Db 2401 AAATTTATTCATGTCCTGCTTCCACGCTGGGGTTTGAATGAAGTACTGGAATGAAGTTTGC 2460
Qy 2461 CTGGATCGCTTGGAAATGCTGAAACCCCACTGCAACCTATCGTTTGGCTCTTGTCAAGA 2520
Db 2461 CTGATCGCTCGGCAATGCTGAAACCACTGCAAGTGTCTTTGGTCTCTATCCAGA 2520
Qy 2521 GGAGGCCATATGTCATGTCGTGCGGGTTAATTAATCAGCCTGCTTTGAACTGTCAAT 2580
Db 2521 GGAGGTCATATGTCCTTGTGTCCTGCTTAAATCAAGCTCTGGAAGTACTGTCAG 2580
Qy 2581 GGGAAACAGAGTGTGAGAGAGAGCTGCGCGCTGTGAGGAGTAGGAAGGAACTTAC 2640
Db 2581 GCGAAACAG 2640
Qy 2641 GGAGTGTCTCGTGGCTGCTGACCAACAGCATCTCTGCTATCATTTCTTTGGCAAAATACT 2700
Db 2641 ACAGTGTCCGAGCGGTGACCGCAACATCTGTTGTCATCATATCTCTGCGCAAAATACT 2700
Qy 2701 TTAATGAGTATGCACTAATGCACTTTTATTTGGTGTATCATATGAAGAGGGTCTCTACAGG 2760
Db 2701 TTAATGAGCATGCACTAATGCACTTTTATTTGGGATCATATGAAGAGGGGCCCCCAGG 2760
Qy 2761 CCACCTTAGACCAAGCAACCCCGAGCTTTCTAAGGCAAGGGGTTCCCTCTCAACTTCCAGT 2820

Qy	2226	CTTCCAAAGTGAAGAAACGATCAAGAGAACATCAAGGAACACCTCCTTGATGATGA	2285
Db	1377	CTTCCAAAGTGAAGAAACGATCAAGAGAACATCAAGGAACACCTCCTTGATGATGA	1436
Qy	2286	AGAGGAGATGAGGAGATATGAGGCGAGAGAAAGGAGAGATGATCTGAATATCGGTC	2345
Db	1437	AGAGGAGATGAGGAGATATGAGGCGAGAGAAAGGAGAGATGATCTGAATATCGGTC	1496
Qy	2346	CAGCAATCAAGCCATGACCCCTATTTAGAAATATATTTAACTATGGACACTGCAGAGCT	2405
Db	1497	CAGCAATCAAGCCATGACCCCTATTTAGAAATATATTTAACTATGGACACTGCAGAGCT	1556
Qy	2406	GTTTATGCTCTGCTTCAAGCCCTGGGGTTTGAATGAAGTACTGGATGAAGTTTGCCTGGA	2465
Db	1557	GTTTATGCTCTGCTTCAAGCCCTGGGGTTTGAATGAAGTACTGGATGAAGTTTGCCTGGA	1616
Qy	2466	TCGCTTTGGAATGCTGAAACCCACATGCAACCGTATGCTTTGGCCCTCTTGTCAAGGAGG	2525
Db	1617	TCGCTTTGGAATGCTGAAACCCACATGCAACCGTATGCTTTGGCCCTCTTGTCAAGGAGG	1676
Qy	2526	CCATATGCTACTGATGCTGCGGGTTTAAATCAGCCTGCTTTGAACTGTACATGGGAA	2585
Db	1677	CCATATGCTACTGATGCTGCGGGTTTAAATCAGCCTGCTTTGAACTGTACATGGGAA	1736
Qy	2586	AACAGAAGTAGGAAGGAGCTGACGCGCTGAGGAGGTAGGAAAGGGAACCTTACGAGT	2645
Db	1737	AACAGAAGTAGGAAGGAGCTGACGCGCTGAGGAGGTAGGAAAGGGAACCTTACGAGT	1796
Qy	2646	GTCGCTGCGCTCACACACAGCATCTCTGTCTATCATTTCTTTGGCAAACTCTTTAAT	2705
Db	1797	GTCGCTGCGCTCACACACAGCATCTCTGTCTATCATTTCTTTGGCAAACTCTTTAAT	1856
Qy	2706	GAGTATGACCAATGCAACTTTTATGGTGATCATATGAAGAAGGTCTTACAGGCCACC	2765
Db	1857	GAGTATGACCAATGCAACTTTTATGGTGATCATATGAAGAAGGTCTTACAGGCCACC	1916
Qy	2766	TAGACCAAGCACCCAGACCTTTCTAAGGCAGAGGGTTCCCTCTCAACTTCCAGTAAATAT	2825
Db	1917	TAGACCAAGCACCCAGACCTTTCTAAGGCAGAGGGTTCCCTCTCAACTTCCAGTAAATAT	1976
Qy	2826	TGTGCAAGGACAGATTAAACAAGTTGCTGCACCTGCTGCTTTCCGCTCGCTGTGATGCTGA	2885
Db	1977	TGTGCAAGGACAGATTAAACAAGTTGCTGCACCTGCTGCTTTCCGCTCGCTGTGATGCTGA	2036
Qy	2886	TCACCTGCGCTCTGACCCCTCTTCTGCTCTGCTCTTACATACCTGTTTCTTAGTAAATGA	2945
Db	2037	TCACCTGCGCTCTGACCCCTCTTCTGCTCTGCTCTTACATACCTGTTTCTTAGTAAATGA	2096
Qy	2946	AGGTTGAGTCAAGTATGCTGATGACATGCTGTTATGCTGCGCAGACCTACTGGGATGGA	3005
Db	2097	AGGTTGAGTCAAGTATGCTGATGACATGCTGTTATGCTGCGCAGACCTACTGGGATGGA	2156
Qy	3006	TAAATTTAGGCTCCCTCTCTGGAGATGCTGGCCCGAAGATGGCAAGATTCGATGCTTGGGA	3065
Db	2157	TAAATTTAGGCTCCCTCTCTGGAGATGCTGGCCCGAAGATGGCAAGATTCGATGCTTGGGA	2216
Qy	3066	GGTGAGAGAACCGGACACAGGCCCTCTTCTGGCGGAACTGAGAGAAATGAGCAGGAGG	3125
Db	2217	GGTGAGAGAACCGGACACAGGCCCTCTTCTGGCGGAACTGAGAGAAATGAGCAGGAGG	2276
Qy	3126	CAGGAAGGAGCCATTTGATGCTGGCTCCTTACTTACCTCAGTACATAGACCGTCAAT	3185
Db	2277	CAGGAAGGAGCCATTTGATGCTGGCTCCTTACTTACCTCAGTACATAGACCGTCAAT	2336
Qy	3186	ATCACCTGGAGTCAATCAGAGCCGCGCAGATATCACACCGGCTCCTGATSCCTCAGG	3245
Db	2337	ATCACCTGGAGTCAATCAGAGCCGCGCAGATATCACACCGGCTCCTGATSCCTCAGG	2396
Qy	3246	GCCTGAAGCAAAAGTCCAGGAGGAGAGCATGACTTGTGATGAGTGAATCACCACTGG	3305
Db	2397	GCCTGAAGCAAAAGTCCAGGAGGAGAGCATGACTTGTGATGAGTGAATCACCACTGG	2456
Qy	3306	TTGCTTATCAAGTGTCCCAAAATGAAAAAATTTTCTACATCTTTACAGGAAGACCGGAA	3366
Db	2457	TTGCTTATCAAGTGTCCCAAAATGAAAAAATTTTCTACATCTTTACAGGAAGACCGGAA	2516
Qy	3366	GCAAGCTACCGCTATTTGTTTACTTTGGAGTAATAGGAGCTGAAATTTGCTGCTGAAATTTGA	3425
Db	2517	GCAAGCTACCGCTATTTGTTTACTTTGGAGTAATAGGAGCTGAAATTTGCTGCTGAAATTTGA	2576
Qy	3426	ACTCTCTAAATCTATTATGACAGACCTCTGAAAGCTCTAGCCAAATTTCTTGAAGGATTTCCGGTT	3485
Db	2577	ACTCTCTAAATCTATTATGACAGACCTCTGAAAGCTCTAGCCAAATTTCTTGAAGGATTTCCGGTT	2636
Qy	3486	GACTAGTGTGATGATCAACTACTGCTGCGCAGACATATCTTTGCAAGGCACTGAGCTTTCT	3545
Db	2637	GACTAGTGTGATGATCAACTACTGCTGCGCAGACATATCTTTGCAAGGCACTGAGCTTTCT	2696
Qy	3546	TCGTCTACAGCCTCAAGCCCCCAACTTCTCCACACAGCACTATCCGAGAAACAGCCAT	3605
Db	2697	TCGTCTACAGCCTCAAGCCCCCAACTTCTCCACACAGCACTATCCGAGAAACAGCCAT	2756
Qy	3606	TGATCTGATTTGACGCTGGGTTCACTGTTTGGGAGCCTTATCATGATGTGTCCGCTGTTCT	3665
Db	2757	TGATCTGATTTGACGCTGGGTTCACTGTTTGGGAGCCTTATCATGATGTGTCCGCTGTTCT	2816
Qy	3666	GATGGGGCTTCTCGAACTTTGTGCGATGTCGCGAGAAACAACTTGCCAACTACAAATGGG	3725
Db	2817	GATGGGGCTTCTCGAACTTTGTGCGATGTCGCGAGAAACAACTTGCCAACTACAAATGGG	2876
Qy	3726	GTTGCTCTGAGCCACAGCAGTCTGCGCCGCTCTGCGAGGATGCGCTCTCGCTCAT	3785
Db	2877	GTTGCTCTGAGCCACAGCAGTCTGCGCCGCTCTGCGAGGATGCGCTCTCGCTCAT	2936
Qy	3786	TGCCACCGCAGACACCCGCTTCTCATCCACCATAGTACCAGAGAGGTACACAGACATAC	3845
Db	2937	TGCCACCGCAGACACCCGCTTCTCATCCACCATAGTACCAGAGAGGTACACAGACATAC	2996
Qy	3846	GGCTCTTGAGCAAAATACCCAAATCAACAGCAAAATATGACACAACTTTGACGAGC	3905
Db	2997	GGCTCTTGAGCAAAATACCCAAATCAACAGCAAAATATGACACAACTTTGACGAGC	3056
Qy	3906	TAAAGGGGAAATTTTGAGAGTCATTGAAATTTCTTATTTGAAAGATGCCACAGATGTTGT	3965
Db	3057	TAAAGGGGAAATTTTGAGAGTCATTGAAATTTCTTATTTGAAAGATGCCACAGATGTTGT	3116
Qy	3966	GGATCTTCTGCTGGAGGTTATGGACATCATTTATGCTGCTTCAAGGATCTTTAGTTAA	4025
Db	3117	GGATCTTCTGCTGGAGGTTATGGACATCATTTATGCTGCTTCAAGGATCTTTAGTTAA	3176
Qy	4026	AAAGAAAGGTCTTCAAGAAATGTTTCCAGCATCTGCAAGGTTCTACATGCTGAGCTATTA	4085
Db	3177	AAAGAAAGGTCTTCAAGAAATGTTTCCAGCATCTGCAAGGTTCTACATGCTGAGCTATTA	3236
Qy	4086	TGAGCGGAATCACAGATAGCAGTTGGAGCTGCG	

Db 3537 CCTCAAGCTGGCCCGGCTCATCTGGACTTCCACCGCAACGTCATCTCATGGCCCATG 3596
|||||
4445 ACAGGAGGAGGACCGCTTCATGCTCTAA 4473
|||||
Db 3597 ACAGGAGGAGGACCGCTTCATGCTCTAA 3625
|||||

RESULT 9
AB011113
LOCUS
DEFINITION Homo sapiens mRNA for KIAA0541 protein, partial cds.
ACCESSION AB011113
VERSION AB011113.1 GI:3043605
KEYWORDS KIAA0541 protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Nagase, T., Ishikawa, K., Miyajima, N., Tanaka, A., Kotani, H.,
Nomura, N., and Ohara, O.
Submitted (13-FEB-1998) Otsu Ohara, Kazusa DNA Research Institute,
DNA Technology, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
[E-mail: cdmainfo@kazusa.or.jp, Tel: +81-438-52-3913,
Fax: +81-438-52-3914]
Location/Qualifiers
1. .6072
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HG4294"
/sex="male"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
1. .6072
/gene="KIAA0541"
1. .3485
/gene="KIAA0541"
/codon_start=3
/product="KIAA0541 protein"
/protein_id="BAA25467.1"
/db_xref="GI:3043606"
/translation="PEFYGCEVFEHLKLLQDSSGELNINWISDTADKQSGEGLANT
TSISLQAEFDKLPAGIIDLQSVIPNSNEPLKVATSVIIPAHGLVCGREDGSIVI
VPATQAI VQLQGEHMLRGPPTLRLGRNKVTCLLYPHQVCSARYDORVLISSGV
DFSVIITDFSGEMHIFCVHGEBITQLLPENPCSRVQHCIVSASDYSVGLLSLR
EKKIMLASRHLFPPIQVIRKPSDDYLVVCGSDGVSVMQMDTGLDRVCVMGTVASLR
LNACDEAPAAVDLSHPAVALKQMTRESLAALKNMHHKQLTLNLLASEADSKG
NLPKYSHNSLQVAIKTNLTDPDIHVLFFDVEALIIQLTEASRPRTALISPNLQK
AGSSDKGSFLTKGAANVLFQVKSTIKENIKHLDDDEEHEIMRORRESPEY
RSKSKPLTLLELNEVDLDEVLDRGLMKPKHTVSLFGLL
LANTLMSMTNATPIGDMKKGPTRPPPTDLSKARGPSPTSSNIVQOIQVAAVP
VSARSADHSNTPGPKLHGKTEVGRKLPASEGVGKGTGVSRAVTOHLLSIIIS
SRGHMSLMIPGNVPACKLSHGTVEGRKLPASEGVGKGTGVSRAVTOHLLSIIIS
VSARSADHSNTPGPKLHGKTEVGRKLPASEGVGKGTGVSRAVTOHLLSIIIS
ARRWQRCLEVRQAQALLAELELRLEQAGKEAIDAWAPYLPQYIDHVSIPGVTSFA
AQYITTPASGPEAKVQEEHLDVDDITTCGLSSVPQMKKISTISEERRKQATV
LGVICAGFEATEPPKLLTRPSSQIPEGFLTSGGNSYLARHTKALFTLLQ
PSPKLPHTIRTAIDLIRGFTWEPYDVSVMGLLELCAEAKOLANI TWGLP
LSPADARSARHALLIATAPPAFTTIAKEVHRHTLAALANTOSQOONHTTLARA
KGEILVRIELIEKMPDVVDLLVEVMDIIMVCLGSLVKKKGLQECFPAICRFYVVS
YVERNHRVAGARHGVSVALYDINTGKQTHIGHKPTITAVAFAPDQRYLATISNTDSH
ISFWQNTSLLSIGMLNSAPQLRCIKTYQVPPQVSPASPGSHNALKARLIIWTSNRNV
ILMAHDGKEHRFMV"

ORIGIN

Query Match 77.9%; Score 3485; DB 9; Length 6072;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 989 CTCGGTCTTCTATGATGACAGAGAAATATTTCCATAAATCTGTTAAATTCAGGGTGAATCTT 1048
Db 1 CTCGGTCTTCTATGATGACAGAGAAATATTTCCATAAATCTGTTAAATTCAGGGTGAATCTT 60
|||||

QY 1049 CTCGAAGTGTGAATATTTTGAACATATCAGACACAGCTGATATAACAGGGAAGTGAAGAAG 1108
Db 61 CTCGAAGTGTGAATATTTTGAACATATCAGACACAGCTGATATAACAGGGAAGTGAAGAAG 120
|||||

QY 1109 GGCTGGCAATGACAACTTCTATTAGTTTGAAGAGGCAATTTGATAAATCTGAATCTCTTGTG 1168
Db 121 GGCTGGCAATGACAACTTCTATTAGTTTGAAGAGGCAATTTGATAAATCTGAATCTCTTGTG 180
|||||

QY 1169 CTGCTGGAATATAGATCAGCTGAGTGTGATTTCCCAATAGTAAATCTCTTAAAGTAA 1228
Db 181 CTGCTGGAATATAGATCAGCTGAGTGTGATTTCCCAATAGTAAATCTCTTAAAGTAA 240
|||||

QY 1229 CTGCAAGTGTGATACATACAGACACATGAGGAGCACTTGTGTTGCTGCTGGAAGATGGAAGCA 1288
Db 241 CTGCAAGTGTGATACATACAGACACATGAGGAGCACTTGTGTTGCTGCTGGAAGATGGAAGCA 300
|||||

QY 1289 TAGTTATTGTACCTGCCACACAGACGCGCCATAGTACAGCTGTTGCAAGGGGAACACATGC 1348
Db 301 TAGTTATTGTACCTGCCACACAGACGCGCCATAGTACAGCTGTTGCAAGGGGAACACATGC 360
|||||

QY 1349 TCAGAAAGAGTGGCCACCTCAGACAACTCCGCTGGTCAATCGGAACAAAGTCAATGTT 1408
Db 361 TCAGAAAGAGTGGCCACCTCAGACAACTCCGCTGGTCAATCGGAACAAAGTCAATGTT 420
|||||

QY 1409 TGCATATCTCATCAGTCTCAGCTCGGTATGATCAAGATACCTGATATCTGGAGGTG 1468
Db 421 TGCATATCTCATCAGTCTCAGCTCGGTATGATCAAGATACCTGATATCTGGAGGTG 480
|||||

QY 1469 TGGATTTTTTCAGTCAATAATTTGGGACATATTTCTCGAGAAATGAAACATATCTTCTGTG 1528
Db 481 TGGATTTTTTCAGTCAATAATTTGGGACATATTTCTCGAGAAATGAAACATATCTTCTGTG 540
|||||

QY 1529 TTCAATGGTGGTGAATATCTCAACTTTCTAGTTCCACTGAAACTCTAGTGCAGAGTAC 1588
Db 541 TTCAATGGTGGTGAATATCTCAACTTTCTAGTTCCACTGAAACTCTAGTGCAGAGTAC 600
|||||

QY 1589 AGCAGTGCATCTGCTGTAGCCAGTGCACACTCAGTGCAGTCTTCTAAGTTTTCGGAGAGA 1648
Db 601 AGCAGTGCATCTGCTGTAGCCAGTGCACACTCAGTGCAGTCTTCTAAGTTTTCGGAGAGA 660
|||||

QY 1649 AAAAATGACATAATGTTGGCATCTCGTCACCTTTTCTTCAAGTAATCAAAATGAGGC 1708
Db 661 AAAAATGACATAATGTTGGCATCTCGTCACCTTTTCTTCAAGTAATCAAAATGAGGC 720
|||||

QY 1709 CTTCTGATGATTAACCTGTGGTGGTGTTCAGATGGTTCGTGTAGTCTGTGCAAAATGG 1768
Db 721 CTTCTGATGATTAACCTGTGGTGGTGTTCAGATGGTTCGTGTAGTCTGTGCAAAATGG 780
|||||

QY 1769 ATACTGGTGCATTTGGATCGTTGTGTGATGGGATTAACAGAGTGTGATTTCTAAACGCTT 1828
Db 781 ATACTGGTGCATTTGGATCGTTGTGTGATGGGATTAACAGAGTGTGATTTCTAAACGCTT 840
|||||

QY 1829 GTGATGAAGCTGTTCTGCTGCTGTTGATTTCACTTAGTCAATCCAGCAGTCAACCTAAAC 1888
Db 841 GTGATGAAGCTGTTCTGCTGCTGTTGATTTCACTTAGTCAATCCAGCAGTCAACCTAAAC 900
|||||

QY 1889 AAGCTATGACGAGAGCTAGTCTTGTGCTCTTAAAAAATATGCCCCATCATTAAGCTACAA 1948
Db 901 AAGCTATGACGAGAGCTAGTCTTGTGCTCTTAAAAAATATGCCCCATCATTAAGCTACAA 960
|||||

QY 1949 CCCTTGGCAACTAACTCTTGGCTCTTGGAGCATCTGCAAGGGAATTTTACCTAAATATT 2008
Db 961 CCCTTGGCAACTAACTCTTGGCTCTTGGAGCATCTGCAAGGGAATTTTACCTAAATATT 1020
|||||

QY	2009	CTCATAACTCCCTGATGGTTCAAGCAATAAAGACAAACCTTAACAGACCCCGGACATACATG	2068	QY	3089	TGCTTCTGCGGAACTGAGAGAAATTGACAGCGCAGGAGGAAGCAATTTGATGCTT	3148
DB	1021	CTCATAACTCCCTGATGGTTCAAGCAATAAAGACAAACCTTAACAGACCCCGGACATACATG	1080	DB	2101	TGCTTCTGCGGAACTGAGAGAAATTGACAGCGCAGGAGGAAGCAATTTGATGCTT	2160
QY	2069	TGCTATTTCTTTGATGTGGAAGCGTTGATTTATTCAACTCTGACTGAAGAAGCCTTAGGC	2128	QY	3149	GGGCTCTTTACTTACCTCTCAGTACATAGACCAAGTCAATATCACTGAGTGCATCATGAAAG	3208
DB	1081	TGCTATTTCTTTGATGTGGAAGCGTTGATTTATTCAACTCTGACTGAAGAAGCCTTAGGC	1140	DB	2161	GGGCTCTTTACTTACCTCTCAGTACATAGACCAAGTCAATATCACTGAGTGCATCATGAAAG	2220
QY	2129	CGAATACCTGCTTTATTTCCCAAGAGAAATTTGCAAAAAGCATCTGGCAGTTTCAGACAAAG	2188	QY	3209	CGCGCAGACATATACACCGGCTCTGATGCTCAGGGGCTGAAAGCAAAAGTCCAGGAGG	3268
DB	1141	CGAATACCTGCTTTATTTCCCAAGAGAAATTTGCAAAAAGCATCTGGCAGTTTCAGACAAAG	1200	DB	2221	CGCGCAGACATATACACCGGCTCTGATGCTCAGGGGCTGAAAGCAAAAGTCCAGGAGG	2280
QY	2189	GGGGCTCTTTTAACTGGAAAAAGAGCAGCAGTCTCTTCCAAACAAAGTGAAGAAGACGA	2248	QY	3269	AAGAGCATCACCTTGTGTGACGATGACATCACCACTGGTGTCTTATCAAGTGTCCCAAAA	3328
DB	1201	GGGGCTCTTTTAACTGGAAAAAGAGCAGCAGTCTCTTCCAAACAAAGTGAAGAAGACGA	1260	DB	2281	AAGAGCATCACCTTGTGTGACGATGACATCACCACTGGTGTCTTATCAAGTGTCCCAAAA	2340
QY	2249	TCAAAGAGACATCAAGGAACCTCTCTGATGATGAAGAGGAGGATGAGGAGATAATGA	2308	QY	3329	TGAAAAAAATTTCTACATCTTACGAGGAAAGACGGAAGCAAGCTACCGCTATTTGTTTTAC	3388
DB	1261	TCAAAGAGACATCAAGGAACCTCTCTGATGATGAAGAGGAGGATGAGGAGATAATGA	1320	DB	2341	TGAAAAAAATTTCTACATCTTACGAGGAAAGACGGAAGCAAGCTACCGCTATTTGTTTTAC	2400
QY	2309	GGCAGAGAGGGAAGAAAGTGATCTCTGAATATCGGTCCAGCAAAATCAAAGCCATTGACCC	2368	QY	3389	TTGGAGTATAGAGCTGAATTTGGTGCTGAAATTTGAACTCTTAACTATTGACAGAC	3448
DB	1321	GGCAGAGAGGGAAGAAAGTGATCTCTGAATATCGGTCCAGCAAAATCAAAGCCATTGACCC	1380	DB	2401	TTGGAGTATAGAGCTGAATTTGGTGCTGAAATTTGAACTCTTAACTATTGACAGAC	2460
QY	2369	TATTAGAAATAATTTAACTATGACACTGCAAAAGCTGTTTATGTCTGCTTCAAGCCCT	2428	QY	3449	CTCGAAGCTCTAGCCAAAATTCCTGAGGGATTCGGGTGAGTACTAGTGTGATCCAACTACT	3508
DB	1381	TATTAGAAATAATTTAACTATGACACTGCAAAAGCTGTTTATGTCTGCTTCAAGCCCT	1440	DB	2461	CTCGAAGCTCTAGCCAAAATTCCTGAGGGATTCGGGTGAGTACTAGTGTGATCCAACTACT	2520
QY	2429	GGGGTTTGAATGAAGTACTGGAATGAAATTTGCTGGATCGCTTGGAACTCTGAAACCC	2488	QY	3509	CGTGGCCAGACATATCTGCAAGGCACTGACGCTTCTCTGTCTACAGCTCCAAAGCCCA	3568
DB	1441	GGGGTTTGAATGAAGTACTGGAATGAAATTTGCTGGATCGCTTGGAACTCTGAAACCC	1500	DB	2521	CGTGGCCAGACATATCTGCAAGGCACTGACGCTTCTCTGTCTACAGCTCCAAAGCCCA	3580
QY	2489	ACTGCAACCGTATCGTTTGGCTCTTGTCAAGAGGAGCCCATATGTCACTGATGCTGCCG	2548	QY	3569	AACTTCTCTCACACAGCATATCCGAGAAACAGCCATTTGATCTGATTTGACGTTGCTCA	3628
DB	1501	ACTGCAACCGTATCGTTTGGCTCTTGTCAAGAGGAGCCCATATGTCACTGATGCTGCCG	1560	DB	2581	AACTTCTCTCACACAGCATATCCGAGAAACAGCCATTTGATCTGATTTGACGTTGCTCA	2640
QY	2549	GTTTAAATCAGCTGCTGTGTAATCTGCATCGGAAACAGAAAGTAGGAGAAAGCTGC	2608	QY	3629	CTGTTTGGAGGCTTACATGGATGTCTCGCTGTTCTGATGGGGCTTCTCGAACTTTGTG	3688
DB	1561	GTTTAAATCAGCTGCTGTGTAATCTGCATCGGAAACAGAAAGTAGGAGAAAGCTGC	1620	DB	2641	CTGTTTGGAGGCTTACATGGATGTCTCGCTGTTCTGATGGGGCTTCTCGAACTTTGTG	2700
QY	2609	CAGCTCTGAGGAGTAGGAAAGGAACTTACCGAGTGCTCCGCTGCGTCACACACAGC	2668	QY	3689	CCGATCCGAGAAACAACTTGGCAACATCAATGAGGGTGTCTCTGAGCCACAGCAGCTG	3748
DB	1621	CAGCTCTGAGGAGTAGGAAAGGAACTTACCGAGTGCTCCGCTGCGTCACACACAGC	1680	DB	2701	CCGATCCGAGAAACAACTTGGCAACATCAATGAGGGTGTCTCTGAGCCACAGCAGCTG	2760
QY	2669	ATCTCTGCTCTATCATTTCTTTGGCAAAATCTTTAATGAGTATGACCAATGCAACTTTTA	2728	QY	3749	ACTCGGCCGCTCTGCGAGGCATGCCCTCTCGCTCATTTGCCACCGCAGACACCCGCT	3808
DB	1681	ATCTCTGCTCTATCATTTCTTTGGCAAAATCTTTAATGAGTATGACCAATGCAACTTTTA	1740	DB	2761	ACTCGGCCGCTCTGCGAGGCATGCCCTCTCGCTCATTTGCCACCGCAGACACCCGCT	2820
QY	2729	TTGGTGAATCATGAGAGGGTCTTACAGGCCACCTAGACCAAGCACCCGAGACCTTT	2788	QY	3809	TCATCACCACTAGCAGAGGTACACAGACATACGGCTCTTTGACAGCAAAATACCCAAT	3868
DB	1741	TTGGTGAATCATGAGAGGGTCTTACAGGCCACCTAGACCAAGCACCCGAGACCTTT	1800	DB	2821	TCATCACCACTAGCAGAGGTACACAGACATACGGCTCTTTGACAGCAAAATACCCAAT	2880
QY	2789	CTAAGGCAAGGGTCTCCCTCCAACTTCCAGTAAATTTGTCAGGAGACAGATTAACAAG	2848	QY	3869	CACAGCAGAATATGCAACAACAACTCTTGCACAGCTTAAAGGGGAAAATTTTCAGAGTCA	3928
DB	1801	CTAAGGCAAGGGTCTCCCTCCAACTTCCAGTAAATTTGTCAGGAGACAGATTAACAAG	1860	DB	2881	CACAGCAGAATATGCAACAACAACTCTTGCACAGCTTAAAGGGGAAAATTTTCAGAGTCA	2940
QY	2849	TTGCTGACCTGCTGTTTCCGCTCGGCTGATGCTGATCTCTGCGCTCTGACCCCTCTT	2908	QY	3929	TTGAAATTTCTTATTTGAAAGATGCCACAGATGTTGAGTCTTCTCGTGAGGTTATGG	3988
DB	1861	TTGCTGACCTGCTGTTTCCGCTCGGCTGATGCTGATCTCTGCGCTCTGACCCCTCTT	1920	DB	2941	TTGAAATTTCTTATTTGAAAGATGCCACAGATGTTGAGTCTTCTCGTGAGGTTATGG	3000
QY	2909	CTGCTCTGCTTTACATACCTGTTTCTTAGTAATGAAAGTTGAGTTCAGTTCAGTCTGCTA	2968	QY	3989	ACATCATTTATGTAAGTCTTGAAGGATCTTTAGTAAAGAAAGAGTCTTCAAGAATGTT	4048
DB	1921	CTGCTCTGCTTTACATACCTGTTTCTTAGTAATGAAAGTTGAGTTCAGTTCAGTCTGCTA	1980	DB	3001	ACATCATTTATGTAAGTCTTGAAGGATCTTTAGTAAAGAAAGAGTCTTCAAGAATGTT	3060
QY	2969	TGCACTGTGTTATGCTGCCAGACCTACTGGGTTGGATAAATTTAGGCTCTCCCTCTCG	3028	QY	4049	TCCAGCCATCTGACGTTCTACATGGTCACTGTTTATGAGCGGAATTCACAGAAATGACAG	4108
DB	1981	TGCACTGTGTTATGCTGCCAGACCTACTGGGTTGGATAAATTTAGGCTCTCCCTCTCG	2040	DB	3061	TCCAGCCATCTGACGTTCTACATGGTCACTGTTTATGAGCGGAATTCACAGAAATGACAG	3120
QY	3029	AGATGCTGGCCCGAAGATGGCAAGATCGATGCTTTGGAGGTGAGAGAAAGCCGACAGGCC	3088	QY	4109	TTGAGCTCGCCATGTTGAGTGGCCCTGTACACATCCGGAATCGGAATGTCAGACAA	4168
DB	2041	AGATGCTGGCCCGAAGATGGCAAGATCGATGCTTTGGAGGTGAGAGAAAGCCGACAGGCC	2100	DB	3121	TTGAGCTCGCCATGTTGAGTGGCCCTGTACACATCCGGAATCGGAATGTCAGACAA	3180
				QY	4169	TCCATGACACAAAGGGAACCAATCACTGAGTGGCTTTTGTCTCTGATGGAAGATATCTTG	4228

|||||
Db 3181 TCCATGGACACAGGACCAATCACTGACGTGGCTTTTTCCTGATGGAAGATATCTTG 3240
Qy 4229 CCACCTACTCAAACTGACAGCCACATTTCTTTTGGCAGATGAACACGTCCTGCTGG 4288
Db 3241 CCACCTACTCAAACTGACAGCCACATTTCTTTTGGCAGATGAACACGTCCTGCTGG 3300
Qy 4289 GAAGCATCGGATGCTGACCTGCGCAGCTTCTGCTGCGTGCATTAACCTACAGGTGC 4348
Db 3301 GAAGCATCGGATGCTGACCTGCGCAGCTTCTGCTGCGTGCATTAACCTACAGGTGC 3360
Qy 4349 CCCCTGTGACCGCGCTCCCGGCTCCCAAGTCCCTCAAGTCCCGCGCTCATCT 4408
Db 3361 CCCCTGTGACCGCGCTCCCGGCTCCCAAGTCCCTCAAGTCCCGCGCTCATCT 3420
Qy 4409 GGACTTCCAAACCGCAAGTCACTCTCATGCGCCATGACGGGAAGGAGCAGCCGCTCATGG 4468
Db 3421 GGACTTCCAAACCGCAAGTCACTCTCATGCGCCATGACGGGAAGGAGCAGCCGCTCATGG 3480
Qy 4469 TCTAA 4473
Db 3481 TCTAA 3485

RESULT 10
AF192379 4371 bp mRNA linear ROD 15-SEP-2001
LOCUS
DEFINITION Rattus norvegicus TGF-beta resistance-associated protein TRAG
ACCESSION AF192379
VERSION AF192379.1 GI:15624076
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 4371)
AUTHORS Sanders, S. and Thorgeirsson, S.S.
TITLE TRAG: a novel gene associated with TGF-beta resistance
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4371)
AUTHORS Sanders, S. and Thorgeirsson, S.S.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1999) LBC, NIH, 37 Convent Drive, Bethesda, MD 20892, USA

FEATURES
source
1..4371
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/cell_type="epithelial"
/tissue_type="liver"
1..>4371
/gene="Trag"
1..>4371
/gene="Trag"
/codon_start=1
/product="TGF-beta resistance-associated protein TRAG"
/protein_id="AA03984.1"
/db_xref="GI:15624076"

translations="MAGNSLVLPVLMGRKAPTHCISILLTDGGTIVTGCHDQIC
LWLSLELYNPRLPFGHTAAITCLSKACASGDKQYTSASANGMCLWDVNDGRCL
EPTKLACTHTGIGIQPQSVGNREGRLCHVYPEILVVDATSLVLSVSKSPDW
ISMSIIRSHRTQEDTVALSVGLKWLVTSEISGLQDTEPIFEESEKPIYCONCO
SLSCFATQBSLLVWCNRYRFDAGDYSLLCSGSDGCTGTGGFVSADKVIWTE
NGQSYITKLSPASCLSDSPRSFSDVGRKAVENLIPVQHSLLDQKRELIVCPVTRFFY
CKEYLLKLDGSSGRSLIWNADTADKOBANEGKLTTCISLQDAFDKLPKCPAG
IIDQLSVIPNSPFLKVTASVYIPAHGRVLCGRDGSIIIVPATQAIVOLLOEHL
RGRWPHTLRHKNVTCLLYPHOVSRDQVLYISGGVDFSVIWIWIFSGMKHIF
CVHGGEITQLLVPPENCARSVQHCVSASDHSVGLLSLEKICMLASRLHPFIQVI
KWPSDYLTVGCTDGSVCVQMDTGALDRCAMGITAVEILINACDEAVPAVDSLHP
AVNLKQAMTTRSALAALKNMAHKLQTLATNLASASDKNLPHYSHNSLMVQAIKTN

LTDPDIHVLFFDVEALLIQLLTBSRPNTALLSPENLQKASSSDKGSGFLTKGRRAA
VLFQVKBTIKENIKHLEDEBEDEBEWREDESPYRASKPFLTLLEYNLATWD
TAKLFKAEKLEHLEDEBEDEBEWREDESPYRASKPFLTLLEYNLATWD
KLJQAKAEAGRGKGPATVTSVAVTQHLISIIISLANLMSMTNATFTGDHM
KKGPTPRPGTDLKARDSPPASSNIQGOIKQWSQLAAHVCMLPDLGLGKFR
PILLEMAREWDRCLVREAOALLLARRIEQGRKETIDTAAPYLQYMDHVLIS
PGTAEMQTMAAAPDASGPEAKVQEBEHLVDITTTGCLSSVPQMKMSSTEEBR
KQATALLGVIGABFGEIPEPKLLTRPRSSQIPGFGLTSGGSNYSLARHTKAL
TFLLOPPSPKLPPTSHTRITADLIGRGTVMPEVDSVGLMLLELCADAEKQAL
NIWGLPLSPADSARSARHSLIATARPAPFTTTAKEVHRHTALAAANTQSOQSH
TTTLARAKGBILRIVIELIKMPTDVVDLLAVEVMDIIMYCLEGLSVKKGLOECPAI
CRFVMSYVERSHRIAVGABHGSVALYDITKQCTIHHGKPIATVSPADGXYLAT
YSNTDHSISFWQNTSLSGISGMSAPQLRCIKTYQVPPVPASPSGNAHLARLI
WTSNRNVILMAHDGKSHRFVW"

ORIGIN

Query Match 75.6%; Score 3382.8; DB 10; Length 4371;
Best Local Similarity 85.5%; Pred. No. 0;
Matches 3820; Conservative 2; Mismatches 549; Indels 99; Gaps 1;
Qy 1 ATGCGAGGAACAGCCCTTGTCTACCCATTTGTTTGGGGTGGAAAGCGCCACACAT 60
Db 1 ATGCGAGGAACAGCTAGTCTGCCCATTGTTCTCTGGGGCGCGCAAGCACACAC 60
Qy 61 TGCATCTCAGCGGTACTTTTAAACAGATGAGGGCCACAGATCGTAAACAGATGTACGAC 120
Db 61 TGTATTTTCGTCAATTAATCTGCTGACAGATGAGGGGGCAGCATTTGTGATGTCAGCAT 120
Qy 121 GGACAAATATGCTCTGGGATCTTTCAGTAGAAGTGCATAATTAATCTCTCAGGACATCTGTG 180
Db 121 GGACAGATATGCTCTGGGATCTCTCAGAAGAGCTGGAAGTTAATCCCCGAGCACTATTA 180
Qy 181 TTTGGTCATACAGCATCAATCACTTTGTTGCTTAAAGCTTGTCTTCCAGTGAACAAACAG 240
Db 181 TTTGGTCATACAGCAGCATCACTTTGTTGCTCAAAAGCCTGTGCTCTGGAGACAAACAG 240
Qy 241 TATATTTGTGATGATCTGAAAGTGGAGATGTGCTCTGGGATGTGATGTGATGGCAGA 300
Db 241 TACACCGTGTAGTGGTCCGCAATGGAGATGTGCTCTGGGATGTGATGTGATGGCAGA 300
Qy 301 TGTATTGAATTTACAAATTTAGCTGTGCACATACATGCGCATACAGTTTCTACCACTTCTCT 360
Db 301 TGTATTGAATTTACAAATTTAGCTGTGCACATACATGCGCATACAGTTTCTACCACTTCTCT 360
Qy 361 GTTGGGAATCAGCGAGAGGAGGCTTTTATGCCACGACATTAACCTCGAAATCTCTTGT 420
Db 361 GTTGGGAATCAGCGAGAGGAGGCTTCTGTGCCATGCGCATTAACCTCGAAATCTCTTGT 420
Qy 421 GTGGATGCTACAGCCTTGAAGTATTATATCTCTTTAGTATCAAGATATCAACAGATGTG 480
Db 421 GTGGATGCCACAGCCTTGAAGTATTATATCTCTTTAGTATCAAGATATCAACAGATGTG 480
Qy 481 ATTAGTCTCATGATGATATTATTCATCCACGACACACAGGACACAGTGTGTAGCACTC 540
Db 481 ATCAGCTCCATGATGATATTATTCATCCACGACACACAGGACACAGTGTGTAGCACTC 540
Qy 541 TCGGTGACTGCGCATCTCGAAGGTCTGGATTGTTTACCTCGGAAATAAGTGCATGACAGAT 600
Db 541 TCGGTGACAGGTATTCTGAAGGTGTGATGTGACCTCTGAAATTAAGTGCATGACAGAT 600
Qy 601 ACTGAGCCAAATATTGAGGAGGAATCCAAACCAATTTATTTGTGATGAGATTCGCCAAAGCATC 660
Db 601 ACTGAGCCAAATATTGAGGAGGAATCCAAACCAATTTATTTGTGATGAGATTCGCCAAAGCATC 660
Qy 661 TCTTTTGTGCAATTTACAAAGGTCACTTTTGGTGTGTGTTCCAAATATTTCGAGGGGTG 720
Db 661 TCTTTTGTGCAATTTACAAAGGTGCTCTTGGTGTGTGTTTCGCAAGTATCGAGGGGTG 720
Qy 721 TTGATGCGCGGAGCATTTCTTGTGTGTGTTTACAGTCTCTAGTCAAGATTCGCAAGATGAG 780
Db 721 TTGATGCGCGGAGCATTTCTTGTGTGTGTTTACAGTCTCTAGTCAAGATTCGCAAGATGAG 780
Qy 781 ACCGGGGGGGAGCATTTTGTCTCATCAGATAAAGTCATCATTTTGGACAGAAAATGGGCAAGT 840

781 ACTGGGGGACTTTGGTCTGCAGACAAAGTCAATATTTGGACTGAAACGGCGCAGAGT 840
841 TATATTTACAACTACCTGCCAGTGGCTTCCAGCTAGTAGTCAATTCACGAGTGTG 900
841 TACATTTACAACTCCCTGCCAGTGGCTTCCAGCTAGTGTATTCATTCGCCAGTGCAGTG 900
901 GGGAGGCGAGTTGAAATTTAATTCCTCCTGTGTACACATATCTCTTGGATCGAAAGAT 960
901 GGGAAAGCAGTTGAAATCTGATTCCTCCGTCGAGCATAGCCCTCTTGGATCGAAGGAT 960
961 AAGAGTTGTCTAATTTGCTCTCTGTGTACTCGGTCTCTATGATGACAGAAATATTC 1020
961 AGAGAGTTGTGAATTTGCTCTCTGTGTACTCGGTCTCTATGATGACAGAAATATTC 1020
1021 CATAAACTGTTAATTCAGGGTGTATTCCTTGGAGGTTGAATATTTGGAAACATATCAGAC 1080
1021 CATAAGCTACTAATTCAGGGTGTATTCCTTGGAGGTTGAATATTTGGAAACATATCAGAC 1080
1081 ACAGCTGATTAACAGGGAGTGAAGAGGGCTGGCAATGACAACTTCTATAGTTTGCAC 1140
1081 ATAGCAGACAAACAGGAAGCCAAATGAAGGGCTTAAACCGAACCTTGTATAGTTTGCAC 1140
1141 GAGGCAATTCATAACTGAATCTCTGTCTCTGTGGAATATAGATCAGCTGAGTGTGAT 1200
1141 GATGCAATTCAGCAACTGAAGCCCTTGTCTGTGGAATATAGATCAGCTGAGTGTGAT 1200
1201 CCCAATAGTAATGAACCTCTTAAAGTAACCTGCAAGTGTGTACATACAGACATGAGACGA 1260
1201 CCAACAGCAACGAACCACTTAAAGTAACTGCGAGTGTGTACATACAGACATGAGACGA 1260
1261 CTGTGTTGGTGTGAGATGAGAGCATAGTATTTGTACTCTGCCACACAGACGGCCATA 1320
1261 CTGTGTTGGCGCGGGAAGA CCGAAGCATCAATATCTGCTGCCACCCAGACGGCCATA 1320
1321 GTACAGCTGTTGCAAGGGGACACATCTCAGAAAGGTTGGCCACCTCACAGAACACTC 1380
1321 GTTCAGCTGTGCGAGGAGAACATCTCTCAGCAGAGTGGCCCCGCGACAGAACCTC 1380
1381 CGTGGCATCGGAA CAAAGTCACATGTTTGTCTATCTCATCAGGTCTCAGCTCGGTAT 1440
1381 CGTGGCCACCGGAACAAAGTCACGTGTTTGTCTATCTCATCAGGTCTCAGCTCGGTAT 1440
1441 GATCAAGATACCTGATATCTGAGGTGTGGATTTTTCAGTCATAATTTGGGACATATTT 1500
1441 GATCAAGATACCTGATATCTGAGGTGTGGATTTTTCAGTCATAATTTGGGACATATTT 1500
1501 TCTGGAGAAATGAACATATCTCTGTGTTTCAATGGTGTGAGATTAATCAACTTCTAGTT 1560
1501 TCCGGAGAAATGAACATATCTCTGTGTTTCAATGGTGTGAGATTAATCAACTTCTAGTT 1560
1561 CCACCTGAAAACCTGTAGTGCAGAGTACAGACATCTGCTCTGTAGCCAGTGACAC 1620
1561 CCGCCAGAAAACCTGTAGTGCAGAGTTCACACTGCTGCTGTTCTGTGGCCAGTGACAC 1620
1621 TCAGTAGACCTTAAAGTTTGGAGAGAAAATGCAATATGTTGGCATCTCGTCACTT 1680
1621 TCTGTAGGCTGTCAAGTCTGCGAGAGAAAATGCAATATGTTGGCGTCTCGTCACTT 1680
1681 TTTCTTATCAAGTAATCAATGAGGCTTCTCATGATTAATCACTGTTGGTGGGTGTTC 1740
1681 TTCTCTATCAGGTGATCAAGTGAAGGCTTCTGACACTACCTGTTGGTGGGTGTCAAG 1740
1741 GATGGTTCTGTGTACGTCTGCAAAATGATATCTGGTCAATTTGGATCTGTGTGATGGG 1800
1741 GACGGCTCTGTGTGTCTGGCAGATGACACACTGGTGGCTGGACCGCTGTGCAATGGG 1800
1801 ATACAGCAGTTGAGATCTTAAACGCTTGTGATGAAGCTGTTCTGTGCTGTGTTGATTC 1860
1801 ATACAGCCCGGAGATTTCTCAATGCTTGTGACGAAGCTGTGCTGTGAGCAGTGGACTCA 1860
1861 CTTAGTCTATCCAGCAGTCAACCTTAAACAAAGCTATGACGACACGTAGTCTTGTGCTCTT 1920

1861 CTCAGTCAACCAGCAGCTCAACCTGAAGCAAGCCATGACACGGCGAGTCTCGCCGCCCTT 1920
1921 AAAAATATGGCCCATCATATAGCTACAAACCTTGGCACTAAACCTCTTGGCTTCTGAGGCA 1980
1921 AAAAATATGGCCCATCAACCAAGCTGCAACCCCTTGGCACTAAACCTTCTTGGCTTCTGAGGCC 1980
1981 TCTGCAAGAGGAAATTTTACATAATATCTCTATAACTCTCTATGATGTTTCAAGCAATAAAG 2040
1981 TCTGCAAGAGGAAATTTTACATAATATCTCTATAACTCTCTATGATGTTTCAAGCAATAAAG 2040
2041 ACAAACTTAACAGACCCGAGCATATCATGTCTATCTTGTGATGGAAGCTTGTATTT 2100
2041 ACAAACTTAACAGACCCGAGCATATCTATGCTCTTCTTGTGATGGAAGCTTGTATTT 2100
2101 CAATCTCTGACTGAAAGAGCCCTAGCCGGAATCTCTTATTTTAACTCGAAAAACGAGCAGCA 2160
2101 CAATCTCTGACTGAAAGAGCCCTAGCCGGAATCTCTATGCTCTTATTTTCCCGAGAGATCTG 2160
2161 CAAAAAGCATCTGGCAGTTTCAAGAAAGGGGCTCTTCTTAACTCGAAAAACGAGCAGCA 2220
2161 CAGAAAGCATCTGGCAGTTTCAAGAAAGGGGCTCTTCTGACTCGAAAAACGAGCAGCA 2220
2221 GTTCTCTTCCAACTGAAAGAAACGATCAAGAGAAACATCAAGGAAACCTCTCTTGAT 2280
2221 GTTCTCTTCCAGCAAGTGAAGAAACATATCAAGAGAAACATCAAGGAGCAGCTCTCTTGAT 2280
2281 GATGAAGAGGAGATGAGAGATTAATGAGCAGAGAGGAGGAGAAAGTATCTGCACTGATAT 2340
2281 GAGGAGGAGGAGCAGAGAGAGGATGAGCAGAGAGGAGGAGAAAGTACCTCTGAGTAC 2340
2341 CGGTCCAGCAAAATCAAAAGCCATTTGACCTATTTAGAAATTAATTTAACTATGACACTGCA 2400
2341 CGGGCCAGCAAGTCCAAACCACTTACCTACTAGATACAACTTACTATGAGATACCGCA 2400
2401 AAGCTGTTATGCTCTGCTCTCAAGCTGGGGTTTGAATGAAGTACTGAGATGAAGTTTGC 2460
2401 AAATTTATCATGCTCTCTCTCAAGCTGGGGTTTGAATGAAGTACTGAGATGAAGTTTGC 2460
2461 CTGATGCTGCTTGGATGCTGAACCCCACTGACCGTATCTGCTCTTCTGCTCTTCTGCAAGA 2520
2461 CTGATGCTGCTTGGATGCTGAACCCCACTGACCGTATCTGCTCTTCTGCTCTTCTGCAAGA 2520
2521 GAGGCGCATATGCTCACTGATGCTGCGGGTTTAAATCAGCTCTCTTGTAACTGTGCACAT 2580
2521 GAGGCTCATATGCTCTTCTGATGCTTCTGTTTAAATCAGGCTCTGGAAGTACTGTCAG 2580
2581 GGGAAACAGAAAGTGAAGAGAACTGCGAGCTCTGAGGGAGTGAAGAAAGGAACTTAC 2640
2581 GCGAAAGCAGAAAGCAGAGCGAAGGGGCCAGCAACGAGAGCGTAGGCAAGGGGACCTAC 2640
2641 GAGTGTCCGTCGCTGCTCACCACAGCATCTCTGCTATCATTTCTTTTGGCAAAATACT 2700
2641 ACAGTGTCCGAGCGGTACACCGCAACATCTGTTGTCATCATATCTCTGGAATAATACT 2700
2701 TTAATGAGTATGACCAATGCACTTTTATTTGTTGATCATATGAAGAGGGTCTCTCACAGG 2760
2701 TTAATGAGTATGACCAATGCACTTTTATTTGTTGATCATATGAAGAGGGTCTCTCACAGG 2760
2761 CCACTTAGACCAAGCAGACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAACTTCCAGT 2820
2761 CCGCTTAGACCAAGCAGACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAGGCTCCAGT 2820
2821 AATATTTGCAAGAGCAGATTAACAAAGTTGCTGCACTCTGCTCTGCTCGCTCGTCTGAT 2880
2821 AACATTTGCAAGAGCAGATTAACAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2845
2881 GCTGATCACTCTGCTCTGACCT 2940
2846 ----- 2845
2941 AATGAAGTTGGAGTCAAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3000
2846 ----AAGGATGAGCCAACTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2901

Qy	3001	TTGGATAAATTTAGGCTCCCTCTTCTGGAGATGCTGGCCCCGAAGATGGCAAGATCGATGC	3060
Db	2902	CTGGGTAAATTTACGGCCCTCTCTTTCTGGAGATGCTAGCTCGAAGATGGCAAGATCGGTGC	2961
Qy	3061	TTGGAGGTGAGAGAAGCGCACAGCCCTGTCTCTGGCGGAACATGAGAAGATTTGAGCAG	3120
Db	2962	TTGGAGGTGAGAGAGGTGACAGCCCTCTTCTTAGCAGAGCTGAGAGAGATTTGAGCAG	3021
Qy	3121	GCAGCGAGGAAGGACCATTTGATGCTCGGCTCTCTTACTTACCTCAGTACATAGACCAC	3180
Db	3022	GCAGCGAGGAGAGACTTATTGATACCTGGCTCTCTTACTTACCTCAGTACATGACCAT	3081
Qy	3181	GTCAATACCTGAGTACATCAGAAGCCGCGCAGACTATCACACGGCTCTGTGATGCC	3240
Db	3082	GTCAATACCTGTGAGTACGCGCGGAGCCATGCGACTATGGCAGCTGTCTCCAGATGCC	3141
Qy	3241	TCAGGCGCTGAAGCAAAAGTCCAGAGGAAGAGCATGCACTTGTGTGACGATGACATCACC	3300
Db	3142	TCGGGCGCAGAGCAAAAGTCCAGAGGAAGAGCATGACCTCTGTGAGAGATGACATCACC	3201
Qy	3301	ACTGGTTGCTTATCAAGTGTCCCAAAATGAAAAAATTTCTACATCTTACGAGGAAAGA	3360
Db	3202	ACTGGTTGCTTATCAAGTGTCCCAAAATGAAAAAGATGTCACATCTTACGAGGAAAGA	3261
Qy	3361	CGGAGCAAGCTACCGCTATTGTTTACTTTGGAGTAAATAGAGCTGAATTTGGTCTGAA	3420
Db	3262	AGGAAGCAGGCCACTGCTATTGTTTCTCTCGGAGTGTAGGAGCAGAGTTTGGAGCTGAA	3321
Qy	3421	ATTGAACCTCTCTAACTATTGACAGACCTCGAAGCTCTAGCCAAAATTCCTGAGGGATTC	3480
Db	3322	ATTGNAACCAAGAAACTGCTGACAGACCTCGAGCTCTAGTCAAAATTCCTGAAGATTT	3381
Qy	3481	GGGTTGACTAGTGGTGGATCCAACTACTCTGCTGCGCAGACATCTTCCAAAGGCATCGACG	3540
Db	3382	GGTTTGACAAAGTGAGGTTCCAACTACTCTGCGCAGACATACGTGCAAGGCATCGACA	3441
Qy	3541	TTTTCTTCTGCTACAGCTCCAGCCCCAAAATTTCTCTCCACAGCACTATCCGAGAACCA	3600
Db	3442	TTTTCTTCTGCTACAGCCCAAGTCCCAAATCTCTCTCATAGCACATCCGAGAACT	3501
Qy	3601	GCCATTGATCTGATTGGACGTGGGTTCACTGTTTGGGAGCTTACATGGATGTGTCCGCT	3660
Db	3502	GCCATTGACCTGATCGGCGAGGGTTCAACGTTGTGGAGCTTACATGGAGCTGTCTGCT	3561
Qy	3661	GTTCGTAGTGGGCTCTTCGAACTTTTGTGCGGATGCGGAGAAAACAACTTGGCCAAATCACA	3720
Db	3562	GTCTGTAGTGGGCTGTCTGGAGCTGTGTGCAGATGCTGAGAAGCAGCTGGGCCAAATCACA	3621
Qy	3721	ATGGGTTGGCTCTGAGCCGACAGCTGACTCGGCCGCTCTGCGGCGATGCCCTCTCG	3780
Db	3622	ATGGGCTGCTCTGAGCCCTGTGAGCTGACTCTGCCCCGATCCGCAAGACAGCCCTTCT	3681
Qy	3781	CTCATTTGCCACCGCAGACCAACCGCTTCACTCACCACTAGTCCAAAGAGGTACACAGA	3840
Db	3682	CTCATAGCCACCGCAGACCAACCGCTTCACTCACCACTAGTCCAAAGAGGTGCGACAGA	3741
Qy	3841	CATACGCTCTTGCAGCAAAATACCCAAATCAGCAGAGATATGCACAAACAACTTTGCA	3900
Db	3742	CACACGGCCCTTGAGCAAAATACCCAGTCCGAGCAGAGTATCCACACCAACCACTGGCA	3801
Qy	3901	CGAGCTAAAGGGGAAATTTTGGAGTCAATTTGAAATCTTATTTGAAAGATGCCACAGAT	3960
Db	3802	AGGCTTAAGGCGGAAATCTTGAGAGTCAATTTGAAATCTTATTCGAAAAGATGSCCTACGGAT	3861
Qy	3961	GTGTGTGATCTTCTCGTGGAGGTTATGGACATCATTTATGTACTGCCCTTGAAGGATCTTTA	4020
Db	3862	GTGTGTGATCTTCTGTGGAGGTCATGACATCATCATGTACTGCTCTGGAGGATCTTTA	3921
Qy	4021	GTTAAAAAGAAAGTCTTCAAGAAATGTTTCCGAGCATCTGCGAGGTTCTATCATGGTGCAGC	4080
Db	3922	GTTAAAGAAAGGGTCTTTCAGGAGTGTTCACAGCACTCTGAGGTTCTATCATGGTGCAGC	3981

Qy	4081	TATTATGAGCGGAATCACAGAAATAGCAGTTGGAGAGCTGCCATGGTTTCAGTGGCCCTGTAC	4140
Db	3982	TATTATGAGCGGAGTCA CAGAAATGCGATTGGAGCAGCCATGGCTCAGTGGCCCTGTAT	4041
Qy	4141	GACATCCGGACTGGAAATGT CAGACAAATCCATGGAACAAGGGACCAATCACTGCAGTGG	4200
Db	4042	GACATCCGGACTGGGAAATGT CAGACAAATCCACGGAACAAAGGGAACCTATCACTGCAGTGG	4101
Qy	4201	GCCTTTTGCTCCTGATGGAAGATATCTTTGCCACCTACTCAAACA CTGACAGCCA CATTTTCT	4260
Db	4102	TCCTTTTGCTCTGATGGCGCTTACCTTTGCCACCTACTCAAACA CTGACAGCCA CATTTTCT	4161
Qy	4261	TTTTTGGCAGATGAACACGTCACCTGCTGGGAAGCATCGGCATGCTGAACTCGGCACCTCAG	4320
Db	4162	TTCTGGCAGATGAACACCTCACCTTCTTGGGAAGCATTTGGCATGTGTAACCTCAGCACCTCAG	4221
Qy	4321	CTGGCTGTGCATTAAACAACCTACCAAGGTGCCCTCTGTCAGCCGCGCTCCCCCGGCTCCAC	4380
Db	4222	CTGGCTGTGCATTAAAGACTTACCAAGTACTTTCAGTGAGCCCGCATCCCTCTGGCTCGCAC	4281
Qy	4381	AATGCCCTCAAGCTTGGCCGGCTCATCTGGAATTTCCAAACCGCAACCTCATCTCATGGCC	4440
Db	4282	AACGCCCTCAGTTGGCCGGCTCATCTGGACTTCCAAACCGGAATGTATCTCTCATGGCC	4341
Qy	4441	CATGACGGGAAGGAGCAACCGCTTCATGGTC	4470
Db	4342	CACGATGGGAAGGAGCAACCGCTTTCATGGTC	4371

RESULT 11

AF188123
 LOCUS 4371 bp mRNA linear ROD 15-SEP-2001
 DEFINITION Mus musculus TGF-beta resistance-associated protein TRAG (Trag)
 mRNA, partial cds.
 ACCESSION AF188123
 VERSION AF188123.1 GI:15624070
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 4371)
 AUTHORS Sanders,S. and Thorgeirsson,S.S.
 TITLE TRAG: a novel gene associated with TGF-beta resistance
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 4371)
 AUTHORS Sanders,S. and Thorgeirsson,S.S.
 TITLE Direct Submission
 JOURNAL Submitted (20-SEP-1999) LEC, NIH, 37 Convent Drive, Bldg. 37, Rm
 3C28, Bethesda, MD 20892-4255, USA
 FEATURES
 source 1..4371
 Location/Qualifiers

gene

CDs

GCKEYJLHLLI QODSSGRLLNIWADI AEQAEDBGL KWTTCTIS QJAPDFKLPJCPAC
I IDQJVSNNPSLEPKAVTVPJPHQSRVY GREDGDSV IIVPATQITV LQJGQEHRL
RRGPHRTLRHNRKNTCLLYPHQSRVYDQRYLIGSGDVF SVIIMV FSGEMKQIT
CVHGSETQLLVPENCSSVQVHCICVSADHSDGIVGIVREKKCCIMASHVAPLSPQIV
KWRPSDYLVSDGDSVYVQWGDALDRCAMGITA VEILNADCAASHVAPLSPQIV
AVNLQKATRSRLAKNNHAKLQTLATNL LASEADQKSNPKYSLNSFWQJAKTN
LTQDPJHLVDFVEALII QLLTAEATLNTL SPENLQKSGSDRGSLFVJTGKRAA
VJFOVKETI KENI KEHLIDDEEBEERARQSDREDPYAKSKPFLTVGVNYQWAG
TAKLFMSCLHAWGLNDEJVC LDRJLGMKPHCTVFCGLSRGCHMSLMLPGVNYQWAG
KLHAKAEVGRKLPAAKGVGKGTVPYTSRAVTTQHLLSI ISLANTLMSWNTAIVTIGDHM
KKGTPRPRGPDLKSAKGVSPASSNIVQGOIKQWGSLOAAHCVMLPDLIGLERPR
PPLLEMLARWQDCELVZEAQALILAE LPRJEAQGEKFTIDTWAKLPJQYMDHVIS
PGVTNEMQTNMAAADPASGEPAKVOBEEDHLDVDDITAGCLSSVVPKMKISITYEHR
KQATAVILVIGAEFAGEEPKLLTPRSSQIPEGFQGVMPDVAVLMLLELADCAAEKQIL
TYLLQPPSKPPHSTIRRTATLDIRGRFTVWPEYDQWDSVAVLMLLELADCAAEKQIL
NITMGLFSPAADSAADHAKLSLATAPPAFTI AEKVEHRTLAALANTQAQSOQSH
TTLTARAKGBIRLIRVILIEKMPDQVDDLVNDEG IIMYCEKGLVTLAAQSGDCFPYAT
CRFTYVSYBERSHIVAGARHGSVALYDRTGCTQTHHCKGSPITAVYSFAPQDGFYAT
YSNTSDSHISFWQNTQJGSLISGMLNASPOLRCICTQYPPVPQPSGSHNALKLARLI
WTSNRSVILMAHDGKGRHFWV

ORIGIN

Query Match	75.3%;	Score 3369.2;	DB 10;	Length 4371;
Best Local Similarity	85.3%;	Pred. No. 0;		
Matches 3813;	Conservative 0;	Mismatches 558;	Indels 99;	Gaps 1;
Qy	1	ATGCGAGGAACAGCCCTGTTTCTACCCATGCTCTTTGGGGTGCAGAAAGCGCCACACAT	60	
Db	1	ATGCGAGGAACAGCCCTGTTTCTGCCCATGTTCTTTGGGGCCGAAAGACCCACACAT	60	
Qy	61	TGCATCTCAGCGGTACTTTTTAACAGATGATGGGGCCACGATCGTTAAACAGAGTGTCA	120	
Db	61	TGCATTTTCGTAACACTCTTGACAGATGATGGGGGCGACAAATTGTAACTGGATGCCAGAT	120	
Qy	121	GGACAAATATGCTCTCGGGATCTTTTTCAGTAGAACTGCAAAATTAATCTCTCGAGCACTGTTG	180	
Db	121	GGACAAATATGCTCTCGGGATGTTTTCGATAGAACTAGAAATTAATCTCCCGAGCACTGTTA	180	
Qy	181	TTTGGTGCATACAGCATCAATCACTGTTTGTCTAAAGCTTGCTCCAGTGACAAACAG	240	
Db	181	TTTGGCCACACAGCATCCATCACATTGTTGTCAAAGCCTCGCTTTGGGGACAGCGG	240	
Qy	241	TATATTGTGATGTCATCTGAAAGTGGAGAGATGTGCCCTCTGGGATGTGAGTGTGGCAGA	300	
Db	241	TACACTGTAGCGCGTCTGCAACCGAGAGATGTGCCCTCTGGGATGTGAACGATGGCAGA	300	
Qy	301	TGTAATGAAATTTACAAATTTAGCTTGACACATCTGGCATACAGTTCTACAGTTCTCT	360	
Db	301	TGTAATGAAATTTACAAATTTAGCCCTGACACACATCTGGCATACAGTTCTACAGTTCTCT	360	
Qy	361	GTTCCGGATCAGCGAGAGGAGGCTTTTATGCCACGGACATTAACCTCGAAATCTCTGTT	420	
Db	361	GTTCCGGAACCGAGAGGAGGCGGCTCTCTGCCATGACATTAACCTCGAAATCTCTGTT	420	
Qy	421	GTGGATGCTACAGCGCTTTGAAGTATTATACCTCTTAGTATCAAAAGATATCAACAGACTGG	480	
Db	421	GTGGATGCCACAGCGCTTGAGTGTGTATTCTCTGGTATCGAAGATCTCTCCAGACTGG	480	
Qy	481	ATTAGTCCCATGATGATTAATTCGATCCCAACCGAACACAAAGAGGACACAGTGGTAGCATC	540	
Db	481	ATCAGCTCCATGAGCATCATCCACTCTCAGCGGACACAAAGAGGACACTGTGGTGGCGCTC	540	
Qy	541	TCGGTGACTGCGATCCTGAAAGTCTGGATGCTTACCTCGGAAATTAAGTGACATCCAGAT	600	
Db	541	TCTGTGACAGGTAATCTGAAGGTGTGGATGTGCACCTCTGAAATTAAGTGGAAATGCAGAT	600	
Qy	601	ACTCAGGCCAATATTGTAGGAGGAATCCAAACCAATTTATTGTCTAGAAATTCGCAAGCATC	660	
Db	601	ACTCAGGCCAATATTGTAGGAGGAATCCAAACCAATTTATTGTCTAGAAATTCGCAAGCATC	660	
Qy	661	TCTTTTTGTGCATTTACACAAAGGTCACTTTTGGTTGTGTTCCTCAAAATATTGGAGGGTG	720	
Db	661	TCTTTTTGTGCATTTACACAGAGTCACTTTTGGTGTGTGTTCCTCAAAATATTGGAGGGTG	720	


```
Db 3862 GTGGTGGATCTTCTTGTGGAGGTGATGGATATCATCATGTACTGCTCGGAAGGATCTTTA 3921
Qy 4021 GTTAAAGAAAGGTCTTCAAGAAATGTTTCCAGCCATCTGCAAGTCTTACATGTCAGC 4080
Db 3922 GTGAAAAAGAGGCTCTTCAGGAGTGTCTTCCAGCTATCTGCAGGTTCTACATGGTCAGC 3981
Qy 4081 TATTATGAGCGGAATCACAGAAATAGCAGTGTGGAGCTCGCCATGTTTCAGTGGCCCTGTAC 4140
Db 3982 TATTATGAGCGCAGTCACAGAAATCGCAGTGTGGAGCAGCCATGGCTCAGTAGCCCTGTAT 4041
Qy 4141 GACATCGGACTCGAAAAATCTCAGACAATCCATGAGCACAAGGAGCAATCACTGCGAGTG 4200
Db 4042 GACATCGGACTCGAAAAATCTCAGACAATCCATGAGCACAAGGAGCAATCACTGCGGCTG 4101
Qy 4201 GCTTTTGTCTCTGATGAAGATATCTTGCACCTACTCAAAACACTGAGAGCCACATTTCT 4260
Db 4102 TCCTTTGTCTCTGATGGCCGTTACCTTGGCCACCTACTCGAACAACCGACGCCACATTTCC 4161
Qy 4261 TTTTGGCAGATGAACAGTCACCTGCTGGGAAGCATCGGCATGCTGAACCTGGCCACCTCAG 4320
Db 4162 TTTGGCAGATGAACACCTCCTCTCTGGGAAGCATTTGAAATGCTGAACCTCCTCCTCAG 4221
Qy 4321 CTGGCTGCATTAACCTACCAAGTCCCTCTGTGAGCCGCGCTCCCGCTCCAC 4380
Db 4222 CTGGCTGCATCAAGACTACCAAGTCCCTCTCAGTGCAGCCAGCATCCCTGCTCCAC 4281
Qy 4381 AATGCCCTCAAGCTGGCCGCTCATCTGACCTTCCAAACCGGAACGTCTCATGCGC 4440
Db 4282 AAGGCCCTTAAGTTGGCCGCTCATCTGACCTTCCAAACCGGAATGTTATCTCTCATGGCC 4341
Qy 4441 CATGACGGGAGGAGCAGCCCTTCATGGTC 4470
Db 4342 CATGATGGGAAGGAACACCGCTTCATGGTC 4371

RESULT 12
AX714633
LOCUS AX714633 3316 bp DNA linear PAT 15-APR-2003
DEFINITION Sequence 1317 from Patent EP1293569.
ACCESSION AX714633
VERSION AX714633.1 GI:29889586
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Isegai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuho, Y.
TITLE Full-length cDNAs
JOURNAL Patent: EP 1293569-A 1317 19-MAR-2003;
Helix Research Institute (JP); Research Association for Biotechnology (JP)
FEATURES
source location/Qualifiers
1. 3316
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 62.8%; Score 2807.4; DB 6; Length 3316;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 3080; Conservative 0; Mismatches 1; Indels 242; Gaps 3;

Qy 1 ATGCGAGGAACAGCCCTTGTCTTACCCATGTTCTTTGGGTCGAAAGCGCCACACAT 60
Db 236 ATGCGAGGAACAGCCCTTGTCTTACCCATGTTCTTTGGGTCGAAAGCGCCACACAT 295
Qy 61 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGCGCACGATCGGTAAACAGGATGTCAGC 120
Db 296 TGCATCTCAGCGGTACTTTTAAACAGATGATGGGCGCACGATCGGTAAACAGGATGTCAGC 355
```

```
Qy 121 GGCAAAATATGTCTCTGGGATCTTTTCAGTAGAACTGCAAAATTAATCTCTGAGCACTGTTG 180
Db 356 GGCAAAATATGTCTCTGGGATCTTTTCAGTAGAACTGCAAAATTAATCTCTGAGCACTGTTG 415
Qy 181 TTTGGTCATACAGCATCAATCTTGTGTTGTCTTAAAGCTTGTGCTTCCAGTGCAAAACAG 240
Db 416 TTTGGTCATACAGCATCAATCTTGTGTTGTCTTAAAGCTTGTGCTTCCAGTGCAAAACAG 475
Qy 241 TATATTTGAGTGATCTGAAAAGTGAGAGATGTGCTCTGGAGTGTGATGTGTCAG 300
Db 476 TATATTTGAGTGATCTGAAAAGTGAGAGATGTGCTCTGGAGTGTGATGTGTCAG 535
Qy 301 TGTATTTGAATTTTACAAAATTTAGCTTGCACACATCTTGGCATACAGTTCTTACAGTTCTCT 360
Db 536 TGTATTTGAATTTTACAAAATTTAGCTTGCACACATCTTGGCATACAGTTCTTACAGTTCTCT 595
Qy 361 GTTGGGAATCAGCGAGGAAGAGGCTTTTATGCCACGGACATTAACCTGAAATCCCTTGT 420
Db 596 GTTGGGAATCAGCGAGGAAGAGGCTTTTATGCCACGGACATTAACCTGAAATCCCTTGT 655
Qy 421 GTGATGCTACAGCTTGAAGTATTTATCTCTTAGTATCAAGATATCACCAGACTGG 480
Db 656 GTGATGCTACAGCTTGAAGTATTTATCTCTTAGTATCAAGATATCACCAGACTGG 715
Qy 481 ATTAGCTCCATGAGTATTTATTCGATCCACCGAAACAAGAGGACACAGTGGTAGCACTC 540
Db 716 ATTAGCTCCATGAGTATTTATTCGATCCACCGAAACAAGAGGACACAGTGGTAGCACTC 775
Qy 541 TCGGTGACCTGCGCATCTGAAAGTCTGGATTGTTTACCTCGGAAATAGTGACATGCAAGAT 600
Db 776 TCGGTGACCTGCGCATCTGAAAGTCTGGATTGTTTACCTCGGAAATAGTGACATGCAAGAT 835
Qy 601 ACTGAGCCAAATTTGAGGAGGAATCCAAACCAATTTATTTGTCAGAAATGCGCAAGCATC 660
Db 836 ACTGAGCCAAATTTGAGGAGGAATCCAAACCAATTTATTTGTCAGAAATGCGCAAGCATC 895
Qy 661 TCTTTTGTGCACTTTTACAAAAGTGCATTTTGTGTTGTTGTTTCCAAATATTTGGAGGGTG 720
Db 896 TCTTTTGTGCACTTTTACAAAAGTGCATTTTGTGTTGTTGTTTCCAAATATTTGGAGGGTG 955
Qy 721 TTCGATGCGGAGACATTTCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 780
Db 956 TTCGATGCGGAGACATTTCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1015
Qy 781 ACCGGGGGAGACTTGTCTCATCAGATAAAGTCATCTTTGGACAGAAAATGGGCAAGT 840
Db 1016 ACCGGGGGAGACTTGTCTCATCAGATAAAGTCATCTTTGGACAGAAAATGGGCAAGT 1075
Qy 841 TATATTTTACAACTACCTCCAGTTGCCCTTCCAGCTAGTGTGATTCATTCGCGAGTGATGTG 900
Db 1076 TATATTTTACAACTACCTCCAGTTGCCCTTCCAGCTAGTGTGATTCATTCGCGAGTGATGTG 1135
Qy 901 GGGAGGCGAGTTGAAAATTTAAATTTCTCTCTGTAACAACATATCTCTTTGGATCGAAAAGAT 960
Db 1136 GGGAGGCGAGTTGAAAATTTAAATTTCTCTCTGTAACAACATATCTCTTTGGATCGAAAAGAT 1195
Qy 961 AAGAGTTGCTAATTTGTTCTCTCTGTTACTCGGTTCTTCTATGATGTCAGAGATATTTTC 1020
Db 1196 AAGAGTTGCTAATTTGTTCTCTCTGTTACTCGG-TCCTTCTATGATGTCAGAGATATTTTC 1254
Qy 1021 CATAAACTGTTAATTCAGGGTGATTTCTTCTGGAAGGTGTAATTTTGGAAACATATCAGAC 1080
Db 1255 CATAAACTGTTAATTCAGGGTGATTTCTTCTGGAAGGTGTAATTTTGGAAACATATCAGAC 1314
Qy 1081 ACAGCTGATAAAACAGGGAAGTGAAGAGGCTGGCAATGACAACTTCTTATTTAGTTGCAA 1140
Db 1315 ACAGCTGATAAAACAGGGAAGTGAAGAGGCTGGCAATGACAACTTCTTATTTAGTTGCAA 1374
Qy 1141 GAGGCAATTTGATTAACCTGAATCCTTGTCTGCTGGAATTTATAGATCAGCTGAGTGTGATT 1200
Db 1375 GAGGCAATTTGATTAACCTGAATCCTTGTCTGCTGGAATTTATAGATCAGCTGAGTGTGATT 1434
```


1201 CCCAATAGTAATGAACCTCTTAAAGTAACCTGCAAGTGTGTACATACCAGCACATGAGCA 1260
1435 CCCAATAGTAATGAACCTCTTAAAGTAACCTGCAAGTGTGTACATACCAGCACATGAGCA 1494
1261 CTTGTTTGTGTCGTGAAGATGGAAGCATAGTTATTTGTAAGTGTGTACCTGCAAGATGAGGCA 1320
1495 CTTGTTTGTGTCGTGAAGATGGAAGCATAGTTATTTGTAAGTGTGTACCTGCAAGATGAGGCA 1554
1321 GTACAGCTGTTGCAAGGGGAACAATGCTCAGAGAGGTTGGCCACCTCAGAGAACATC 1380
1555 GTACAGCTGTTGCAAGGGGAACAATGCTCAGAGAGGTTGGCCACCTCAGAGAACATC 1614
1381 CGTGGTCATCGGAACCAAGTCAATGTTGCTATATCTCATCAGTCTCAGCTCGTAT 1440
1615 CGTGGTCATCGGAACCAAGTCAATGTTGCTATATCTCATCAGTCTCAGCTCG --- 1670
1441 GATCAAGATACCTGATATCTGAGAGTGTGGATTTTTCAGTCATATTTGGGACATATTT 1500
1671 ----- 1670
1501 TCTGGAGAAATGAACAATATCTTCTGTGTTTCATGGTGGTGAGATTACTCAACTTCTAGTT 1560
1671 ----- 1670
1561 CCACCTGAACCTGTAGTGAAGATACAGCATGCACTGCTCTGTAGCCAGTACCAC 1620
1671 -----GCAAGATACAGCATGCACTGCTCTGTAGCCAGTACCAC 1712
1621 TCAGTAGGACTTCTTAAGTTTGCAGAGAAATAATGCAATATTTGGCATCTCGTCACTT 1680
1713 TCAGTAGGACTTCTTAAGTTTGCAGAGAAATAATGCAATATTTGGCATCTCGTCACTT 1772
1681 TTTCTCTATTCAAGTAATCAAAATGGAGCCCTTCTGATGATTAACCTGGTGGTGGGTTCTCA 1740
1773 TTTCTCTATTCAAGTAATCAAAATGGAGCCCTTCTGATGATTAACCTGGTGGTGGGTTCTCA 1832
1741 GATGGTCTGTGTACGTCGCAATGGATATCTGGTGCATTTGGATGATGATGATGATGGG 1800
1833 GATGGTCTGTGTACGTCGCAATGGATATCTGGTGCATTTGGATGATGATGATGATGGG 1892
1801 ATAAACAGCAGTGTGAGATTCTAAACGCTTGTGATGAAGCTGTCTGCTGCTGTGATTCA 1860
1893 ATAAACAGCAGTGTGAGATTCTAAACGCTTGTGATGAAGCTGTCTGCTGCTGTGATTCA 1952
1861 CTTAGTCATCAGCAGTCAACCTAAACAAAGCTATGACGAGAGTGTCTGCTGCTCTT 1920
1953 CTTAGTCATCAGCAGTCAACCTAAACAAAGCTATGACGAGAGTGTCTGCTGCTCTT 2012
1921 AAAAATATGGCCCATCAATAGCTACAAACCTTTGCAACTAACTCTTGGCTTCTGAGGCA 1980
2013 AAAAATATGGCCCATCAATAGCTACAAACCTTTGCAACTAACTCTTGGCTTCTGAGGCA 2072
1981 TCTGACAAGGGAATTTTACCTAAATATCTCAATACCTCCTGATGTTCAAGCAATAAG 2040
2073 TCTGACAAGGGAATTTTACCTAAATATCTCAATACCTCCTGATGTTCAAGCAATAAG 2132
2041 ACAAACTTAACAGACCCGACATACATGCTATTTCTTGAATGGAAGCGTTGATTAAT 2100
2133 ACAAACTTAACAGACCCGACATACATGCTATTTCTTGAATGGAAGCGTTGATTAAT 2192
2101 CAATCTCTGATGAGAGGCTCTAGCCGAATATCTGCTCTTAATTTCCCGCAGAGATTG 2160
2193 CAATCTCTGATGAGAGGCTCTAGCCGAATATCTGCTCTTAATTTCCCGCAGAGATTG 2252
2161 CAAAAGCATCTGCGAGTTGAGCAAGGGGCTCTTTTAACTGGAACCAAGCAGCA 2220
2253 CAAAAGCATCTGCGAGTTGAGCAAGGGGCTCTTTTAACTGGAACCAAGCAGCA 2312
2221 GTTCTCTTCCAAAGTAAGAAACGATCAAGAGAAATCAAGGAACACCTCTCTTGTAT 2280
2313 GTTCTCTTCCAAAGTAAGAAACGATCAAGAGAAATCAAGGAACACCTCTCTTGTAT 2372
2281 GATGAAGAGAGGATGAGGAGATAATGAGGAGAGGGAAGAAAGTGTATCTGTAATAT 2340

RESULT 13

2373 GATGAAGAGAGGATGAGGAGATAATGAGCAGAGAGGGAAGAAAGTGTATCTGTAATAT 2432
2341 CGGTCAGCAAAATCAAAAGCCATTGACCTATTAGAATATAATTTAACTATGGACACTGCA 2400
2433 CGGTCAGCAAAATCAAAAGCCATTGACCTATTAGAATATAATTTAACTATGGACACTGCA 2492
2401 AAGCTGTTATGCTCTGCTTCAAGCTGCGGGTTTGAATGAAGTACTGGAATGAAGTTTGC 2460
2493 AAGCTGTTATGCTCTGCTTCAAGCTGCGGGTTTGAATGAAGTACTGGAATGAAGTTTGC 2552
2461 CTGGATGCGCTTGGAAATGCTGAAACCCCACTGCAAGTATCTGTTGGCCTCTTGTCAAGA 2520
2553 CTGGATGCGCTTGGAAATGCTGAAACCCCACTGCAAGTATCTGTTGGCCTCTTGTCAAGA 2612
2521 GGAAGGCCATATGCTGATGCTGCGGGTTTAAATCAGCTGCTGTTGAACTGTACAT 2580
2613 GGAAGGCCATATGCTGATGCTGCGGGTTTAAATCAGCTGCTGTTGAACTGTACAT 2672
2581 GGGAAAACAGAGATGAGAAAGGAGCTGCCAGCTCTGAGGAGTAGGAAAGGAACTTAC 2640
2673 GGGAAAACAGAGATGAGAAAGGAGCTGCCAGCTCTGAGGAGTAGGAAAGGAACTTAC 2732
2641 GGAAGTCTCGGTCGCTGCTGCAAGCATCTCTGCTATCATTTCTTTGGCAATACT 2700
2733 GGAAGTCTCGGTCGCTGCTGCAAGCATCTCTGCTATCATTTCTTTGGCAATACT 2792
2701 TTAATCAGTATGACCAATGCAACTTTTATTTGGTGATCATATGAAGAGGCTCTTACCAGG 2760
2793 TTAATCAGTATGACCAATGCAACTTTTATTTGGTGATCATATGAAGAGGCTCTTACCAGG 2852
2761 CCACCTAGACCAAGCACCCAGACCTTTCTAAGGCAAGGGTTCCCTCCCACTTCCAGT 2820
2853 CCACCTAGACCAAGCACCCAGACCTTTCTAAGGCAAGGGTTCCCTCCCACTTCCAGT 2912
2821 AATATTTGTCAGAGGACAGATTAAACAAGTTGCTGACCTGCTGCTTCCGCTCGTCTGAT 2880
2913 AATATTTGTCAGAGGACAGATTAAACAAGTTGCTGACCTGCTGCTTCCGCTCGTCTGAT 2937
2881 GCTGATCACTCTGGCTCTGACCCCTCTCTGCTGCTCTTTTACATACCTGTTCTTATGTA 2940
2938 ----- 2937
2941 AATGAAGTTGGAGTCAGTTAGTGTCTATGCTATGCTGTGTTATGCTGCCAGACCTTACGGGA 3000
2938 ---AAGGTTGGAGTCAGTTAGTGTCTATGCTATGCTGTGTTATGCTGCCAGACCTTACGGGA 2993
3001 TTGGATAAATTTAGGCTCTCCCTTCTGGAGATGCTGCGCCGGAAGATGCGAAGATCGATGC 3060
2994 TTGGATAAATTTAGGCTCTCCCTTCTGGAGATGCTGCGCCGGAAGATGCGAAGATCGATGC 3053
3061 TTGGAGGTGAGAGAGCCGACAGGCTCTGCTGCGGAACCTGAGAGAAATTTGAGCAG 3120
3054 TTGGAGGTGAGAGAGCCGACAGGCTCTGCTGCGGAACCTGAGAGAAATTTGAGCAG 3113
3121 GCAGGCAAGAGAGCCCAATGATGCTGCGGCTCTTACTTACTCTCAGTACATAGACCAC 3180
3114 GCAGGCAAGAGAGCCCAATGATGCTGCGGCTCTTACTTACTCTCAGTACATAGACCAC 3173
3181 GTCATATCACTGGAGTCATCATCAGAGCCGCGCAGATCATCAACAGGCTCTGATGCC 3240
3174 GTCATATCACTGGAGTCATCATCAGAGCCGCGCAGATCATCAACAGGCTCTGATGCC 3233
3241 TCAGGSCCTGAAGCAAAAGTCCAGGAGGAGAGATGACCTTGTTCACATGATCATCACC 3300
3234 TCAGGSCCTGAAGCAAAAGTCCAGGAGGAGAGATGACCTTGTTCACATGATCATCACC 3293
3301 ACTGGTTGCTTATCAAGTGTCCC 3323
3294 ACTGGTTGCTTATCAAGTGTCCC 3316

AK057141	3316 bp	mRNA	linear	PRI 30-JAN-2004
LOCUS				
DEFINITION	Homo sapiens cDNA FLJ32579 fis, clone SPLEN2000267, highly similar to Rattus norvegicus TGF-beta resistance-associated protein (TRAG) mRNA.			
ACCESSION	AK057141			
VERSION	AK057141.1 GI:16552730			
WORDS	oligo capping; fis (full insert sequence).			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1			
AUTHORS	Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Oabayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, S., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hirao, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, H., Tanase, I., Nomura, Y., Togiya, S., Konai, P., Hara, R., Hara, A., Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Ichihara, T., Sasaki, N., Aotsuka, S., Yoshioka, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hiki, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Wakabashi, Y., Kikuchi, H., Nakagawa, K., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, K., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isegai, T. and Sugano, S.			
TITLE	Complete sequencing and characterization of 21,243 full-length human cDNAs			
JOURNAL	Nat. Genet. 36 (1), 40-45 (2004)			
PUBLISHED	14702039			
REFERENCE	2			
AUTHORS	Kawakami, B., Sugiyama, A., Takemoto, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Iehii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isegai, T.			
TITLE	NEDO human cDNA sequencing project			
JOURNAL	Unpublished			
REFERENCE	3 (bases 1 to 3316)			
AUTHORS	Isegai, T., Otsuki, T. and Sugiyama, T.			
TITLE	Direct Submission			
JOURNAL	Submitted (24-OCT-2001) Takao Isegai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)			
COMMENT	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.			
FEATURES	Location/Qualifiers			
source	1. .3316 /organism="Homo sapiens"			

Query Match

Best Local Similarity

Mismatches

Conservative

0

Score

2807.4

DB

9

Length

3316

Matches

3080

Indels

242

Gaps

3

1

ATGGCAGGAAACAGCCTTGTCTACCCATTGTTCTTTGGGGTCGAAAAGCGCCACACAT

60

236

ATGGCAGGAAACAGCCTTGTCTACCCATTGTTCTTTGGGGTCGAAAAGCGCCACACAT

295

61

TGCATCTCAGCGGTACTTTTAAACAGATGATGGGCGCACGATCGTAAACAGGATGTCAAGAC

120

296

TGCATCTCAGCGGTACTTTTAAACAGATGATGGGCGCACGATCGTAAACAGGATGTCAAGAC

355

121

GGACAAATATGTCTCTGGATCTTTTCAGTAGAAGTGCCTTAAATTAATCTCGACATCTTGT

180

356

GGACAAATATGTCTCTGGATCTTTTCAGTAGAAGTGCCTTAAATTAATCTCGACATCTTGT

415

181

TTTGGTCTACAGCATCAATCACTTGTCTTAAAGCTTGTGCTTCCAGTGCACAAACAG

240

416

TTTGGTCTACAGCATCAATCACTTGTCTTAAAGCTTGTGCTTCCAGTGCACAAACAG

475

241

TATATTGTGAGTCATCTGAAAGTGGAGAGATGTGCTCTGGATGTGAGTGAATGGCAGA

300

476

TATATTGTGAGTCATCTGAAAGTGGAGAGATGTGCTCTGGATGTGAGTGAATGGCAGA

535

301

TGTATTGAATTTACAAATTTAGTTCACACATATCGCATACAGTCTTACCAGTTCTCT

360

536

TGTATTGAATTTACAAATTTAGTTCACACATATCGCATACAGTCTTACCAGTTCTCT

595

361

GTGGGAATCAGCAGAGAGAGGCTTTTATGCCAGGACATATACCTCGAAATCTCTTGT

420

596

GTGGGAATCAGCAGAGAGAGGCTTTTATGCCAGGACATATACCTCGAAATCTCTTGT

655

421

GTGGATGTACAGCCTTGAAGTATATATCTCTTAGTATCAAGATATCAACAGCTGG

480

656

GTGGATGTACAGCCTTGAAGTATATATCTCTTAGTATCAAGATATCAACAGCTGG

715

481

ATTAGTCCATCAGTATTTATTCGATCCACCGAACACAGAGGACACAGTGTGACATC

540

716

ATTAGTCCATCAGTATTTATTCGATCCACCGAACACAGAGGACACAGTGTGACATC

775

541

TCGGTGACTGGCATCTCGAAGGCTCTGGATTGTTTACCTCGAAATAGTGACATGCAAGAT

600

776

TCGGTGACTGGCATCTCGAAGGCTCTGGATTGTTTACCTCGAATTAAGTGACATGCAAGAT

835

601

ACTGAGCCAAATATTTGAGGAGGAATCCAAACCAATTTATGTGAGATTTGCCAAGCATC

660

836

ACTGAGCCAAATATTTGAGGAGGAATCCAAACCAATTTATGTGAGATTTGCCAAGCATC

895

661

TCCTTTTGTGCACTTTACAAAGGTCATTTTGTGTTGTTGTTCCAAATATTTGGAGGTTG

720

896

TCCTTTTGTGCACTTTACAAAGGTCATTTTGTGTTGTTGTTCCAAATATTTGGAGGTTG

955

721

TTTCGATGCGGAGACATTTCTTGTGTTGTTTCCAGTCTTAGTCAAAATGGACAGACATGG

780

956

TTTCGATGCGGAGACATTTCTTGTGTTGTTTCCAGTCTTAGTCAAAATGGACAGACATGG

1015

781

ACCGGGGGGAGCTTTGTCTCATCAGATAAAGTCATCATTTGGACAGAAAATGGGCAAGT

840

1016

ACCGGGGGGAGCTTTGTCTCATCAGATAAAGTCATCATTTGGACAGAAAATGGGCAAGT

1075

841

TATATTACAAACTACCTGCCAGTTCCTTCCAGCTAGTGATTTCATTCGCGAGTGTATGG

900

1076

TATATTACAAACTACCTGCCAGTTCCTTCCAGCTAGTGATTTCATTCGCGAGTGTATGG

1135

901

GGGAAGGCAGTTTGAAAAATTTAAATTCCTCTGTACAAATATCTCTTGGATCGAAAAGAT

960

1136 GGGAGGCGAGTTGAAATTTAAATTCCTCTGTACAACATATCTCTTGGATCGAAAGAT 1195
961 AAGAGGTTGCTAAATTTGCTCTCTGTTACTCGGTTCTTCTATGATGACAGAAATATTC 1020
1196 AAGAGGTTGCTAAATTTGCTCTCTGTTACTCGG-TCTTCTATGATGACAGAAATATTC 1254
1021 CATAACTGTTAAATTCAGGAGTGAATCTCTGGAAGGTTGAATATTTGGAACATATCAGAC 1080
1255 CATAACTGTTAAATTCAGGAGTGAATCTCTGGAAGGTTGAATATTTGGAACATATCAGAC 1314
1081 ACAGCTGATAAACAGGGAAGTGAAGAGGCTGCAATGACAACTTCTATAGTTTGGCAA 1140
1315 ACAGCTGATAAACAGGGAAGTGAAGAGGCTGCAATGACAACTTCTATAGTTTGGCAA 1374
1141 GAGGCAATTTGATAAACTGAACTCTTGTCTCTGCTGGAATTAAGATCAGCTGAGTGTGAT 1200
1375 GAGGCAATTTGATAAACTGAACTCTTGTCTCTGCTGGAATTAAGATCAGCTGAGTGTGAT 1434
1201 CCCAATAGTAAATGAACCTCTTAAAGTAACTGCAAGTGTGTACATACAGACATGAGACGA 1260
1435 CCCAATAGTAAATGAACCTCTTAAAGTAACTGCAAGTGTGTACATACAGACATGAGACGA 1494
1261 CTTGTTTGTGCTGCTGAAGTGAAGCATAGTATTATGTAACCTGCCACACAGACGGCCATA 1320
1495 CTTGTTTGTGCTGCTGAAGTGAAGCATAGTATTATGTAACCTGCCACACAGACGGCCATA 1554
1321 GTACAGCTGTTGCAAGGGGAACACATGCTCAGAAAGGTTGGCCACCTCACAAGAACATC 1380
1555 GTACAGCTGTTGCAAGGGGAACACATGCTCAGAAAGGTTGGCCACCTCACAAGAACATC 1614
1381 CGTGGTCATCGGAACAAAGTCAATGTTTGTCTATATCTCATCAGTCTCAGCTCGGTAT 1440
1615 CGTGGTCATCGGAACAAAGTCAATGTTTGTCTATATCTCATCAGTCTCAGCTCG- 1670
1441 GATCAAAGATACCTGATATCTGAGGTGTGGAATTTTTCAGTCATAATTTGGGACATATT 1500
1671 - 1670
1501 TCTGGAGAAATGAAACATATCTTCTGTGTTTCATGGTGTGAGATTACTCAACTCTAGTT 1560
1671 - 1670
1561 CCACCTGAAACCTGTAGTGAAGAGTACAGCATGCACTGCTCTGTAGCCAGTACAC 1620
1671 - GCAAGATACAGCATGCACTGCTCTGTAGCCAGTACAC 1712
1621 TCAGTAGGACTTCAAGTTTGCAGAGAAATGCAATATGTTGGCATCTGCTCACCTT 1680
1713 TCAGTAGGACTTCAAGTTTGCAGAGAAATGCAATATGTTGGCATCTGCTCACCTT 1772
1681 TTTCTTATTCAGATTAATCAAAATGGAGGCTTCTGATGATTAACCTGGTGGTGGGTGTTCA 1740
1773 TTTCTTATTCAGATTAATCAAAATGGAGGCTTCTGATGATTAACCTGGTGGTGGGTGTTCA 1832
1741 GATGGTCTGTGATCGTGTGGGAAATGGATATGTTGGTGCATTTGGATCGTTGTGATGGG 1800
1833 GATGGTCTGTGATCGTGTGGGAAATGGATATGTTGGTGCATTTGGATCGTTGTGATGGG 1892
1801 ATACACAGCTTTGAGATTCTAAACGCTTGTGATGAAGCTGTTCTGCTGCTGTTGATTCA 1860
1893 ATACACAGCTTTGAGATTCTAAACGCTTGTGATGAAGCTGTTCTGCTGCTGTTGATTCA 1952
1861 CTTAGTCATCCAGCAGTCAACTAAACAGCTATGACAGAGCTAGTCTTGTGCTCTT 1920
1953 CTTAGTCATCCAGCAGTCAACTAAACAGCTATGACAGAGCTAGTCTTGTGCTCTT 2012
1921 AAAAATATGCCCATCATAGCTACAAACCTTTGCAACTAACTCTTGGCTTCTGAGGCA 1980
2013 AAAAATATGCCCATCATAGCTACAAACCTTTGCAACTAACTCTTGGCTTCTGAGGCA 2072
1981 TCTGACAGGGAAATTTACCTAAATATTTCTCAATCTCCCTGATGTTTCAAGCAATAAAG 2040

2073 TCTGACAGGGAAATTTACTTAATATTTCTCATACCTCCCTGATGTTCAAGCAATAAAG 2132
2041 ACAAACTTAACAGACCCGACATACATGTGCTATTCTTTGATGTGAAGCGTTGATATT 2100
2133 ACAAACTTAACAGACCCGACATACATGTGCTATTCTTTGATGTGAAGCGTTGATATT 2192
2101 CAACTCTGACCTGAAGAGGCTCTAGGCGGATACCTCTTATTTTCCCGAGAGATTTG 2160
2193 CAACTCTGACCTGAAGAGGCTCTAGGCGGATACCTCTTATTTTCCCGAGAGATTTG 2252
2161 CAAAAGCATCTGGCAGTTTCAAGAAAGGGGCTCTTTTAACTCGGMAAACGAGACGA 2220
2253 CAAAAGCATCTGGCAGTTTCAAGAAAGGGGCTCTTTTAACTCGGMAAACGAGACGA 2312
2221 GTTCTCTTCCAACTGAAGAAACGATCAAGAGAAACATCAAGGAACACCTCTTGTAT 2280
2313 GTTCTCTTCCAACTGAAGAAACGATCAAGAGAAACATCAAGGAACACCTCTTGTAT 2372
2281 GATGAAGAGGAGGATGAGGAGATTAATGAGCAGAGAGGAGGAGAAAGTGTCTGATAT 2340
2373 GATGAAGAGGAGGATGAGGAGATTAATGAGCAGAGAGGAGGAGAAAGTGTCTGATAT 2432
2341 CGGTCCAGCAAAATCAAGGCATTTGACCTATTAGATAATAATTTAACTATGGAACATGCA 2400
2433 CGGTCCAGCAAAATCAAGGCATTTGACCTATTAGATAATAATTTAACTATGGAACATGCA 2492
2401 AAGCTGTTTATGCTGCTCTTCAAGCTGGGGTTTGAATGAAGTACTGGAATGAAGTTTC 2460
2493 AAGCTGTTTATGCTGCTCTTCAAGCTGGGGTTTGAATGAAGTACTGGAATGAAGTTTC 2552
2461 CTGGATCGCTTGGAAATGCTGAAACCCCACTGCACCGTATCGTTTGGCTCTTGTGACGA 2520
2553 CTGGATCGCTTGGAAATGCTGAAACCCCACTGCACCGTATCGTTTGGCTCTTGTGACGA 2612
2521 SGAGGCCATATGCTGATGCTGCGGGTTAATCAAGCTGCTGTTGTAACCTGTCACAT 2580
2613 SGAGGCCATATGCTGATGCTGCGGGTTAATCAAGCTGCTGTTGTAACCTGTCACAT 2672
2581 GGGAAACAGAAAGTAGAAAGGAGCTGCCAGCGTCTGAGGGAGTAGAAAGGAACTTAC 2640
2673 GGGAAACAGAAAGTAGAAAGGAGCTGCCAGCGTCTGAGGGAGTAGAAAGGAACTTAC 2732
2641 SGAGTGTCCGTGCGCTGACCAACAGCATCTCTGCTATCATTTCTTTTGGCAAAATACT 2700
2733 SGAGTGTCCGTGCGCTGACCAACAGCATCTCTGCTATCATTTCTTTTGGCAAAATACT 2792
2701 TTAATGAGTATGACCAATGCAATTTTATTTGGTGTATCATATGAAGAGGCTCTACAGG 2760
2793 TTAATGAGTATGACCAATGCAATTTTATTTGGTGTATCATATGAAGAGGCTCTACAGG 2852
2761 CCACCTTAGAACCAAGCACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAACTTCCAGT 2820
2853 CCACCTTAGAACCAAGCACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAACTTCCAGT 2912
2821 AATATTGTCAAAGGACAGATTAAACAAAGTTGTGTGACCTGTGCTGTTTCCGCTCGGTCTGAT 2880
2913 AATATTGTCAAAGGACAGATTAAAC- 2937
2881 GCTGATCACTCTGGCTCTGACCCCTCTCTGCTCTCTTACATACCTGTTTCTTAGTA 2940
2938 - 2937
2941 AATGAAGTTGGAGTCAGTTAGCTGTATGCTATGCTGTGTTATGCTGCAGACCTTACTGGGA 3000
2938 - - - - - AAGTTTGGAGTCAGTTAGCTGTATGCTGTGTTATGCTGCAGACCTTACTGGGA 2993
3001 TTGGAATAATTTAGGCTCTTCTTGGAGATGCTGGCCGGAAGATGGAAGATCGATGC 3060
2994 TTGGAATAATTTAGGCTCTTCTTGGAGATGCTGGCCGGAAGATGGAAGATCGATGC 3053
3061 TTGGAAGTGTAGAGAGCCGACAGACCCCTGCTTCTGGCGGAATGTAGAGAAATTTAGCAG 3120
3054 TTGGAAGTGTAGAGAGCCGACAGACCCCTGCTTCTGGCGGAATGTAGAGAAATTTAGCAG 3113

QY 1654 TGCATATGTTGGCATCTCGTACCTTTTTCCTTATTCAGATTAATCAAAATGGAGGCGCTTCT 1713
Db 781 TGCATCATGTTGGCGTCTCGCCACCTCTTCTTATTCAGGTGATCAAGTGGAGGCGCTTCC 840
QY 1714 GATGATTACCTTGGTGGGCTGTTCCAGATGGTTCCTGATGACCTCTGGCAAAATGGATACT 1773
Db 841 GACGATACCTTGGTGGGATGACAGAGCGGCTCTGTGTATGTTCTGGCAGATGGACACT 900
QY 1774 GGTGCATTTGGATCGTTGTGTGATGGGATACAGCAGTTGAGATTTCTAAACGCTTTGTGAT 1833
Db 901 GGTGGCGTGCATCGTTGTGCAATGGGAATCACAGCAGTGGAGATCTGNAAGCTTTGTGAC 960
QY 1834 GAAGCTGTTCTGCTGCTGTTGATTCACTTAGTATCATCCAGCAGTCAACCTTAAACAGCT 1893
Db 961 GAAGCGGTTCTGCGCGTGTAGACTCTCTTAGTATCATCCAGCAGTCAACCTTGAACAGGCC 1020
QY 1894 ATGACGAGAGCTAGTCTGTGCTCTTAAATAATATGCCCCATCATAGCTTACAAACCCCTT 1953
Db 1021 ATGACAGAGCGAGTCTGCGCGCTCTTAAATAATATGCCCCCACCATAAGCTTGAACCCCTT 1080
QY 1954 GCAACTAACCTCTTGGCTCTGAGGCACTCTGACAAAGGAAATTTACCTAAATATTTCTCAT 2013
Db 1081 GCAACGAACCTTTTGGCTCTGAGGCTCTGACAGGGTAAATTTACCTAAATATTTCCCAT 1140
QY 2014 AACTCCCTGATGTTCAAGCAATAAAGACAAACCTTAACAGCCCGACATACATGTGCTA 2073
Db 1141 AACTCCCTGATGTTCAAGCAATAAAGACAAACCTTAACAGCCCGGATATCCATGTGCTT 1200
QY 2074 TTCTTTGATGAGAGGTTGATTTTCACTCTGACTGACAGAGGCTCTAGGCGCAAT 2133
Db 1201 TTCTTTGATGAGAGGTTTGTATTTTCACTCTGACTGACAGAGGCTCTAGGCGCAAT 1260
QY 2134 ACTGCTCTTATTTCCCGAGAGATTTGCAAAAAGCATCTGGCAGTTTCAGACAAAGGGGC 2193
Db 1261 ACTGCACTTATTTCCCGAGAGATCTGACAAAGCATCTGGCAGTTTCAGACAAAGGGGC 1320
QY 2194 TCTTTTAACTGGAAAAAGCAGCAGTCTCTTCCAAACAGTGAAGAAACGATCAAA 2253
Db 1321 TCTTTTCTTAACCTGGGAAGCAGCAGCAGTCTCTTTCCAGCAAGTGAAGAAACATATCAAA 1380
QY 2254 GAGAACTCAAGAAACACCTCTTGTATGATGAAGAGGAGATGAGGAGATATAGAGGCAG 2313
Db 1381 GAGAACTAAGAAACACCTCTTGGATGAGAGGAGGAGCAGAGAGGAGGAGGAGCAG 1440
QY 2314 AGAAGGAGAAAGTGAATCTGTAATTCGGTCCAGCAAAATCAAGCCATTTGACCCCTATTA 2373
Db 1441 AGCAGGAGGATAGTGAACCTTGATACCGGGCCAGCAAGTCCAAAGCACTGACCCCTACTA 1500
QY 2374 GAATATAATTTAACTATGGCACTGCAAAAGCTGTTATGTCTGCTTCAAGCTTGGGGT 2433
Db 1501 GAATACAACTTACTATGGATACCGCAAAATTTATCATGTCTCTGCTCCAGGCTGGGGC 1560
QY 2434 TTGAATGAAGTACTGATGAAGTTTGGCTGGATGCGCTTGGATGCTGGAATCTGAAACCCCACTGC 2493
Db 1561 TTGAATGAAGTCTGATGAAGTTTGTCTGATGCGCTCGGAATGCTGAAACCGCACTGC 1620
QY 2494 ACCGATCGTTTGGCTCTTGTCAAGAGGAGGCATATGTCACTGATGCTGCCGGGTAT 2553
Db 1621 ACCGATCGTTTGGCTCTTGTCAAGAGGAGGTATATGTGCGTGTGATGCTTCCGGGTAT 1680
QY 2554 AATCAGCCTGCTTTGTAACCTGTCAATGGGAAACAGAAATGAGAAAGAGTGGCCAGCG 2613
Db 1681 AATCAGCCTGCTGTAACCTGTGCAATGGGAAACAGAAATGAGAAAGAGTGGCCAGCA 1740
QY 2614 TCTGAGGAGTGAAGAAAGGAACTTACGAGTGTCCGCTGCGGTGACCAAGCAAGCATCTC 2673
Db 1741 GCGGAGGCGGTGAGCAAGGAACTTACAGTGTCTCGAGCGGTCAACCTCAGCATCTG 1800
QY 2674 CTGCTATCATTTCTTTGGCAATATCTTTAAGTATGACCAATGCACTTTTATGCT 2733
Db 1801 TTGTTCATCATATCCCTTGGCGAAATCTTTAATGATATGAACCAATGCACTTTTATGGC 1860

QY 2734 GATCATATGAAGAGGCTCTTACCAGGCCACCTAGACCAAGACCCACAGACCTTTCTTAAG 2793
Db 1861 GATCATATGAAGAGGCGCCACACAGGCGCCTTAGACCTGGCCACCCACAGACCTTCTTAAG 1920
QY 2794 GCAAGGGTTCCTCCCTCCAACTTCCAGTAAATTTGTCAGAGGACAGATTAACAAAGTTGCT 2853
Db 1921 GCGAGGATTTCCCTCCAGCTCCAGTAAATTTGTGCAAGGACAGATTAAC----- 1972
QY 2854 GCACCTGTGCTTTCCGCTCGGTCTGATGTGATCACTCTGGCTCTGACCCCTCTTCTGCT 2913
Db 1973 ----- 1972
QY 2914 CCGTCTTACATACCTGTTTCTTAGTAAATGAAGGTTGGAGTCAGTTAGCTGTGATGACAC 2973
Db 1973 -----AAGGCTGGAGCCAGTTAGCGGCATGACAC 2001
QY 2974 TGTGTTATGCTGCGACAGCTTACTGGGATTTGATTAATTTAGGCTCCCTTTCTGGAGATG 3033
Db 2002 TGTGTTATGTTGCTGACCTGCTGGGCTGGAGATTTAGGCTCTCTCTCTGGAGATG 2061
QY 3034 CTGGCCCGAAGATGGCAAGATCGATGCTTTGGAGTGAAGAACCGCACAGGCGCTTCT 3093
Db 2062 CTAGCTCGAAGATGGCAGGACCGATGCTTTGGAGTGAAGAGGCTGCAAGGCGCTTCT 2121
QY 3094 CTGGCCGAACTGAGAAAGATTTGAGCAGGAGGAGGAGGAAAGCAATTTGATGCTGGCT 3153
Db 2122 CTAGCTGAGCTGAGAAAGATTTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2181
QY 3154 CTTTACTTACCTCAGTACATAGACCACTCATATCACTGGAGTCAATCAGAGGCGCG 3213
Db 2182 CTTTACTTACCTCAGTATATGAGACCATGTCATATCACTGGGCTCAAGCGGAGGAGGAG 2241
QY 3214 CAGACTATCACCGGCTCTGATGCTCAGGCGCTTGAAGAAAGTTCAGGAGGAGGAG 3273
Db 2242 CAGACTATGCGAGCTGCTCCAGTGTCTGGGCGGAGAGGAGGAGGAGGAGGAGGAG 2301
QY 3274 CATGACTTGTGTCAGATGACATCACCACTGTTGCTTATCAAGTGTCCACAAATGAAA 3333
Db 2302 CATGACTGTTGATGATGACATCACTGAGGTTGCTTGTCAAGTGTCCACAGATGAAA 2361
QY 3334 AAAATTTCTACATCTTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3393
Db 2362 AAAATCTCCAGCTTTACGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2421
QY 3394 GTAATAGGAGCTGAATTTGGTGTCTGAAATTTGAACCTCTTAAACTATTTGACCAAGCTCGA 3453
Db 2422 GTGATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2481
QY 3454 AGCTTAGCCAAATTCCTGAGGAGTTGGGTTGACTGATGTTGGTGGATCCAACTACTCGCTG 3513
Db 2482 AGCTCCAGTCAAAATCCCTGAAGGCTTTGGCTTTGACCGAGGAGGCTCCAATCTCCCTG 2541
QY 3514 GCGAGACATCTTGCAGGCACTGAGCTTTCTTCTGCTACAGCTTCAAGCCCAAACTT 3573
Db 2542 GCGAGACACACTGTGCAAGGCACTGACATATCTTCTGCTACAGCCCAAGTCCCAAGCTC 2601
QY 3574 CTTCCACACAGCACTATCCGAAGAACAGCAATTTGATTTGGAGTGGGTTTCACTGTT 3633
Db 2602 CTTCTCATAGCAACATCCGAGGAACTGCACTGACTGATTTGGGAGGAGGTTCACTGTA 2661
QY 3634 TGGAGGCTTACATGATGTGTCGCTGTTCTGATGGGCTTCTCGAACTTTGTGCCGAT 3693
Db 2662 TGGAGGCTTACATGAGCGTGTCCGCTGCTGATGGGCTGCTGGAGCTCTGTGAGAC 2721
QY 3694 GCGGAGAAACATTTGCCAAACATCAATGGGCTTCTGAGGCCAGGAGTGAATCG 3753
Db 2722 GCTGAGAAACATTTGCCAAACATCAATGGGCTTCTGAGGCCCTGAGGCTGAGTCA 2781
QY 3754 GCGGCTCTGCGAGGAGTGGCTTCTGCTCATTTGCGCCAGGAGGAGGAGGAGGAGGAGGAG 3813
Db 2782 GCGGCTCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2841
QY 3814 ACCACCATAGCCAAAGAGGTACACAGACATACGGCTCTTTGAGCAAAATATACCAATCACAG 3873

Query Match		41.28;	Score 1841.8;	DB 9;	Length 4460;
Best Local Similarity		99.94;	Pred. No. 0;		
Matches 1843;		Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	2629	AAGGGAACCTTACCGAGTGTCCCGTCCGCTCAACACACAGCATCTCTCTGTCTATCATTTCT	2688		
DB	1	AAGGGAACCTTACCGAGTGTCCCGTCCGCTCAACACACAGCATCTCTCTGTCTATCATTTCT	60		
QY	2689	TTGGCAAAATCTTTAATGAGTATGACCAATGCAACTTTTATTTGGTGTATATGAAGAAG	2748		
DB	61	TTGGCAAAATCTTTAATGAGTATGACCAATGCAACTTTTATTTGGTGTATATGAAGAAG	120		
QY	2749	GGTCTTACCAAGGCCACTTAGACCAAGCACCCGAGACCTTTCTAAGGCAAGGGGTCCCT	2808		
DB	121	GGTCTTACCAAGGCCACTTAGACCAAGCACCCGAGACCTTTCTAAGGCAAGGGGTTCCT	180		
QY	2809	CCAACTTCCAGTAAATATTTGCAAGGACAGATTAAACAAGTTGCTGCACCTGTCTTTC	2868		
DB	181	CCAACTTCCAGTAAATATTTGCAAGGACAGATTAAACAAGTTGCTGCACCTGTCTTTC	240		
QY	2869	GCTCGGTCTGATCTGATCACTCTGGCTCTGACCCCTCTTCTGCTCTCTTTACATACC	2928		
DB	241	GCTCGGTCTGATCTGATCACTCTGGCTCTGACCCCTCTTCTGCTCTCTTTACATACC	300		
QY	2929	TGTTTCTTAGTAAATGAAGTTTGGAGTTCAGTTAGCTGCTATGCACTGTGTATGCTGCCA	2988		
DB	301	TGTTTCTTAGTAAATGAAGTTTGGAGTTCAGTTAGCTGCTATGCACTGTGTATGCTGCCA	360		
QY	2989	GACCTACTGGATTGGATAAATTTAGGCTTCCCTCTCTGGAGATGCTGGCCCGAAGATGG	3048		
DB	361	GACCTACTGGATTGGATAAATTTAGGCTTCCCTCTCTGGAGATGCTGGCCCGAAGATGG	420		
QY	3049	CAAGATCGATGCTTGGAGGTGAGAGAAGCCGACAGGCCCTCTGCTTCTGGCGGAACCTGAGA	3108		
DB	421	CAAGATCGATGCTTGGAGGTGAGAGAAGCCGACAGGCCCTCTGCTTCTGGCGGAACCTGAGA	480		
QY	3109	AGAATTGAGCAGGACGAGGAAGGAAGCAATGATGCTGGCTCTTACTTACTCTCAG	3168		
DB	481	AGAATTGAGCAGGACGAGGAAGGAAGCAATGATGCTGGCTCTTACTTACTCTCAG	540		
QY	3169	TACATAGACCACTCATATCACCTGGAGTCACATCAGAACCGCGCAGACTATCACACG	3228		
DB	541	TACATAGACCACTCATATCACCTGGAGTCACATCAGAACCGCGCAGACTATCACACG	600		
QY	3229	GCTCCTGATGCTCAGGCTTGAAGCAAAAGTCCAGGAGGAAGCATGACCTTGTGAC	3288		
DB	601	GCTCCTGATGCTCAGGCTTGAAGCAAAAGTCCAGGAGGAAGCATGACCTTGTGAC	660		
QY	3289	GATGACATCACCACTGTGTCTTATCAAGTGTCCCAAAATGAAAAATTTCTACATCT	3348		
DB	661	GATGACATCACCACTGTGTCTTATCAAGTGTCCCAAAATGAAAAATTTCTACATCT	720		
QY	3349	TACGAGGAAGACGGAAGCAAGCTACCGCTATTGTTTACTTGGAGTAATAGAGCTGAA	3408		
DB	721	TACGAGGAAGACGGAAGCAAGCTACCGCTATTGTTTACTTGGAGTAATAGAGCTGAA	780		
QY	3409	TTTGGTGTGAAATTTGAACTCTTAACTATTGACACAGACTCTGAAAGCTCTAGCCAAAT	3468		
DB	781	TTTGGTGTGAAATTTGAACTCTTAACTATTGACACAGACTCTGAAAGCTCTAGCCAAAT	840		
QY	3469	CCTGAGGATTCGGGTGACTAGTGTGGATCAACTACTCTGCTGGCCAGACATCTTGC	3528		
DB	841	CCTGAGGATTCGGGTGACTAGTGTGGATCAACTACTCTGCTGGCCAGACATCTTGC	900		
QY	3529	AAGGCACTGACGTTTCTTCTGCTACAGCTCCAGGCCCAAACTTCTCTCACAACAGCACT	3588		
DB	901	AAGGCACTGACGTTTCTTCTGCTACAGCTCCAGGCCCAAACTTCTCTCACAACAGCACT	960		
QY	3589	ATCCGAAGAACAGCATTGATCTGATTTGACGTGGGTTCATCTGTTGGGAGCCTTACATG	3648		
DB	961	ATCCGAAGAACAGCATTGATCTGATTTGACGTGGGTTCATCTGTTGGGAGCCTTACATG	1020		
QY	3649	GATGTGTGCGCTGTTCTGTGATGGGGCTTCTCGAACTTTGTGCCGATCCGAGAAACAACCT	3708		

Search completed: July 2, 2005, 09:20:50
Job time : 18391 secs

DB	1021	GATGTGTGCGCTGTCTGTGATGGGGCTTCTGAACTTTGTGCCGATCCGAGAAACAACCTT	1080
QY	3709	GCCAAACATCACAAATGGGGTTGCTCTGTAGCCACAGAGCTGACTCGGCCCGCTCTGGGAGG	3768
DB	1081	GCCAAACATCACAAATGGGGTTGCTCTGTAGCCACAGAGCTGACTCGGCCCGCTCTGGGAGG	1140
QY	3769	CATGCCCTCTCGCTCATTTGCCACCGCCAGACCAACCGCTTATCATCAACCATAGCCAAA	3828
DB	1141	CATGCCCTCTCGCTCATTTGCCACCGCCAGACCAACCGCTTATCATCAACCATAGCCAAA	1200
QY	3829	GAGGTACACAGACATACGGCTCTTTCAGCAAAATACCAATACAGCAGATATGCAACA	3888
DB	1201	GAGGTACACAGACATACGGCTCTTTCAGCAAAATACCAATACAGCAGATATGCAACA	1260
QY	3889	ACAACTTTCGACGAGCTAAAGGGGAAATTTTGAGAGTCAATGAAATTTCTATTGAAAAAG	3948
DB	1261	ACAACTTTCGACGAGCTAAAGGGGAAATTTTGAGAGTCAATGAAATTTCTATTGAAAAAG	1320
QY	3949	ATGCCACACAGATGTTGTGATCTTCTCTGGAGAGTTATGGAATCATATATGTAATGCTT	4008
DB	1321	ATGCCACACAGATGTTGTGATCTTCTCTGGAGAGTTATGGAATCATATATGTAATGCTT	1380
QY	4009	GAGGATCTTTAGTTTAAAGAAAGGTCTTCAAGATGTTTCCAGCCATCTGCAGGTTTC	4068
DB	1381	GAGGATCTTTAGTTTAAAGAAAGGTCTTCAAGATGTTTCCAGCCATCTGCAGGTTTC	1440
QY	4069	TACATGCTCAGCTATTATGAGCGGAATCACAGAAATAGCAGTTGGAGCTGCCATGTTCA	4128
DB	1441	TACATGCTCAGCTATTATGAGCGGAATCACAGAAATAGCAGTTGGAGCTGCCATGTTCA	1500
QY	4129	GTGGCCCTGTACGACATCCGGAATGTGAAATGTGAGCAATCCATGGACACAGGAGCCA	4188
DB	1501	GTGGCCCTGTACGACATCCGGAATGTGAAATGTGAGCAATCCATGGACACAGGAGCCA	1560
QY	4189	ATCAGTGCAGTGGCTTTGCTCTGTATGGAAGATATCTTGCACCTACTCTCAACACTGAC	4248
DB	1561	ATCAGTGCAGTGGCTTTGCTCTGTATGGAAGATATCTTGCACCTACTCTCAACACTGAC	1620
QY	4249	AGCCACATTTCTTTTGGCAGATGAACAGTCACTGCTGGGAAGCATCGGCATGCTGAAC	4308
DB	1621	AGCCACATTTCTTTTGGCAGATGAACAGTCACTGCTGGGAAGCATCGGCATGCTGAAC	1680
QY	4309	TCGGCACCTCAGCTGGCTGCATTTAAACCTACAGGTGCCCTCTGTGAGCCCGCTTC	4368
DB	1681	TCGGCACCTCAGCTGGCTGCATTTAAACCTACAGGTGCCCTCTGTGAGCCCGCTTC	1740
QY	4369	CCGGCTCCCAATGTCCTCAAGCTGGCCCGGCTCATCTGGAATTTCCAAACCGCAAGTC	4428
DB	1741	CCGGCTCCCAATGTCCTCAAGCTGGCCCGGCTCATCTGGAATTTCCAAACCGCAAGTC	1800
QY	4429	ATCCTCATGCGCCCATGACGGGAAGGAGCACCGCTTCTCATGCTCTAA 4473	
DB	1801	ATCCTCATGCGCCCATGACGGGAAGGAGCACCGCTTCTCATGCTCTAA 1845	

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 03:39:26 ; Search time 2385 seconds
(without alignments)
11763.304 Million cell updates/sec

Title: US-10-645-335-1
Perfect score: 4473
Sequence: 1 atggcaggaaacagccttgt.....agcaccgtctatggtctaa 4473

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 6313374 seqs, 3136092125 residues

Total number of hits satisfying chosen parameters: 12626748

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4473	100.0	4473	21	US-10-645-335-1 Sequence 1, Appli
2	4460.2	99.7	4517	17	US-10-093-463-153 Sequence 153, Appl
3	3497	78.2	6224	17	US-10-120-988-91 Sequence 91, Appl
4	2807.4	62.8	3316	17	US-10-094-749-1317 Sequence 1317, Ap
5	474.6	10.6	2578	17	US-10-108-260A-1051 Sequence 1051, Ap
6	150.2	3.4	585	16	US-10-029-386-7862 Sequence 7862, Ap
7	147.2	3.3	227	16	US-10-029-386-21562 Sequence 21562, A

8	125.4	2.8	585	13	US-10-027-632-96119 Sequence 96119, A
9	125.4	2.8	585	17	US-10-027-632-96119 Sequence 96119, A
10	125.4	2.8	590	13	US-10-027-632-320926 Sequence 320926, A
11	125.4	2.8	590	17	US-10-027-632-320926 Sequence 320926, A
12	125.4	2.8	598	13	US-10-027-632-237407 Sequence 237407, A
13	125.4	2.8	598	17	US-10-027-632-237407 Sequence 237407, A
14	125.4	2.8	598	13	US-10-027-632-237408 Sequence 237408, A
15	125.4	2.8	598	17	US-10-027-632-237408 Sequence 237408, A
16	103.8	2.3	501	16	US-10-029-386-8422 Sequence 8422, Ap
17	99	2.2	99	16	US-10-029-386-8422 Sequence 8422, Ap
18	88.4	2.0	555	16	US-10-029-386-8040 Sequence 8040, Ap
19	84	1.9	84	16	US-10-029-386-21740 Sequence 21740, A
20	57	1.3	65	10	US-09-908-975-3606 Sequence 3606, Ap
21	57	1.3	572	13	US-10-027-632-192215 Sequence 192215, A
22	57	1.3	572	17	US-10-027-632-192215 Sequence 192215, A
23	41.8	0.9	852	17	US-10-369-493-42703 Sequence 42703, A
24	41.4	0.9	786	19	US-10-437-963-19710 Sequence 19710, A
25	41.2	0.9	1007	20	US-10-425-115-1650 Sequence 1650, Ap
26	40.8	0.9	1159	17	US-10-398-221-1331 Sequence 1331, Ap
27	40.8	0.9	1995	17	US-10-398-221-3359 Sequence 3359, Ap
28	40.6	0.9	774	18	US-10-287-971-71 Sequence 71, Appl
29	39.6	0.9	514	16	US-10-029-386-4824 Sequence 4824, Ap
30	39.6	0.9	678	18	US-10-425-114-10913 Sequence 10913, A
31	39.6	0.9	701	18	US-10-424-599-89082 Sequence 89082, A
32	39.6	0.9	861	18	US-10-424-599-89104 Sequence 89104, A
33	39.6	0.9	1076	18	US-10-424-599-130835 Sequence 130835, A
34	39.6	0.9	312477	19	US-10-317-883A-12 Sequence 12, Appl
35	39	0.9	177556	10	US-09-952-213D-6 Sequence 6, Appl
36	38.8	0.9	1583	18	US-10-425-114-2214 Sequence 2214, Ap
37	38.8	0.9	2762	20	US-10-425-115-162201 Sequence 162201, A
38	38.8	0.9	67883	22	US-10-737-082-66 Sequence 66, Appl
39	38.8	0.9	67883	22	US-10-765-790-66 Sequence 66, Appl
40	38.6	0.9	230	9	US-09-867-701-9505 Sequence 9505, Ap
41	38.4	0.9	1694	20	US-10-425-115-38361 Sequence 38361, A
42	38.2	0.9	6482	8	US-08-781-986A-52 Sequence 52, Appl
43	38.2	0.9	6482	18	US-10-329-624-52 Sequence 52, Appl
44	38	0.8	902	18	US-10-424-599-59957 Sequence 59957, A
45	38	0.8	1170	17	US-10-282-122A-34230 Sequence 34230, A

ALIGNMENTS

RESULT 1
US-10-645-335-1
; Sequence 1, Application US/10645335
; Publication No. US20050064430A1
; GENERAL INFORMATION:
; APPLICANT: Masakazu Takeuchi
; APPLICANT: Yoshiaki Takai
; TITLE OF INVENTION: Rabconnectin-3-binding protein
; FILE REFERENCE: 2003946-0055
; CURRENT APPLICATION NUMBER: US/10/645,335
; CURRENT FILING DATE: 2003-08-21
; PRIOR APPLICATION NUMBER: 2002-319521
; PRIOR FILING DATE: 2002-11-01
; NUMBER OF SEQ ID NOS: 4
; SEQ ID NO 1
; LENGTH: 4473
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(4470)
US-10-645-335-1

Query Match 100.0%; Score 4473; DB 21; Length 4473;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCAGGAAACAGCCTTGTCTTACCCATTGTTCTTTGGGTGCGAAAGCGCCACAT 60
DB 1 ATGGCAGGAAACAGCCTTGTCTTACCCATTGTTCTTTGGGTGCGAAAGCGCCACAT 60

QY	61	TCGATCTCAGCGTACTTTTAA	CAGATGATGGGGCCACGATCGTAA	CAGATGTCACGAC	120	
Db	61	TCGATCTCAGCGTACTTTTAA	CAGATGATGGGGCCACGATCGTAA	CAGATGTCACGAC	120	
QY	121	GGACAAATATGTCTCTGGGATCTTT	CAGTAGAACTGCAAAATTAATCTCT	CAGACATGTTG	180	
Db	121	GGACAAATATGTCTCTGGGATCTTT	CAGTAGAACTGCAAAATTAATCTCT	CAGACATGTTG	180	
QY	181	TTTGGTCATACAGCATCAATCA	CTTGTTGCTTAAAGCTTGCTCCAGTGA	CAAAACAG	240	
Db	181	TTTGGTCATACAGCATCAATCA	CTTGTTGCTTAAAGCTTGCTCCAGTGA	CAAAACAG	240	
QY	241	TATATTGTAGTCATCTGAACTGAA	GTGGAGATGCTCTGGATGTGATGGCAGA		300	
Db	241	TATATTGTAGTCATCTGAACTGAA	GTGGAGATGCTCTGGATGTGATGGCAGA		300	
QY	301	TGTATTGAAATTTACAAATTTAG	CTTGCCACATACCTGGCATACAGTTCT	ACCAAGTTCTCT	360	
Db	301	TGTATTGAAATTTACAAATTTAG	CTTGCCACATACCTGGCATACAGTTCT	ACCAAGTTCTCT	360	
QY	361	GTTGGGAATCAGCGAGAGAAAG	CTTTTATGCGACATTTAGCCAGCA	ATTACCTGTTGTT	420	
Db	361	GTTGGGAATCAGCGAGAGAAAG	CTTTTATGCGACATTTAGCCAGCA	ATTACCTGTTGTT	420	
QY	421	GTGGATGCTACCAAGCTTGAAG	TATATCTCTTAGTATCAAGATATC	ACCAGACTGG	480	
Db	421	GTGGATGCTACCAAGCTTGAAG	TATATCTCTTAGTATCAAGATATC	ACCAGACTGG	480	
QY	481	ATTAGCTCCATGATGATTTAT	TTCGATCCACCGCAACACAGAG	CACAGTGGTAGCACTC	540	
Db	481	ATTAGCTCCATGATGATTTAT	TTCGATCCACCGCAACACAGAG	CACAGTGGTAGCACTC	540	
QY	541	TCGGTGAATGGCATCTGAAAG	CTGGAATTTGTTACCTCGGAA	ATAAGTGAATGCAATG	600	
Db	541	TCGGTGAATGGCATCTGAAAG	CTGGAATTTGTTACCTCGGAA	ATAAGTGAATGCAATG	600	
QY	601	ACTGAGCCAAATTTGAGGAG	GAATCCAAACCAATTTATTTG	TGTCAGAAATGCCAAACATC	660	
Db	601	ACTGAGCCAAATTTGAGGAG	GAATCCAAACCAATTTATTTG	TGTCAGAAATGCCAAACATC	660	
QY	661	TCCTTTTGTGATTTACAAAG	CTCACTTTGTTGTTGTTCCAA	ATAATTTGGAGGTTG	720	
Db	661	TCCTTTTGTGATTTACAAAG	CTCACTTTGTTGTTGTTCCAA	ATAATTTGGAGGTTG	720	
QY	721	TTCCGATCCCGAGACTAT	CTTGTGTTCTAGTCTAGTGAA	AAATGACAGACATGG	780	
Db	721	TTCCGATCCCGAGACTAT	CTTGTGTTCTAGTCTAGTGAA	AAATGACAGACATGG	780	
QY	781	ACGGGGGGGACTTTGTTCT	CATCAGATAAAGTCAATTTGG	ACAGAAATGGCCAAAGT	840	
Db	781	ACGGGGGGGACTTTGTTCT	CATCAGATAAAGTCAATTTGG	ACAGAAATGGCCAAAGT	840	
QY	841	TATATTACAACTACCTCGC	AGTTGCTTCCAGCTAGTGAT	CAATTTCCCGAGTGATG	900	
Db	841	TATATTACAACTACCTCGC	AGTTGCTTCCAGCTAGTGAT	CAATTTCCCGAGTGATG	900	
QY	901	GGGAAGGCGTGTGAAATTT	TAATTCCTCTGTACAAATAT	TCCTTTGGATCGAAAGAT	960	
Db	901	GGGAAGGCGTGTGAAATTT	TAATTCCTCTGTACAAATAT	TCCTTTGGATCGAAAGAT	960	
QY	961	AAAGAGTTGTAATTTGCT	CTCTGTTTACTCGGTTCTTCT	TATGGATCGAGAAATATTC	1020	
Db	961	AAAGAGTTGTAATTTGCT	CTCTGTTTACTCGGTTCTTCT	TATGGATCGAGAAATATTC	1020	
QY	1021	CATAAACTGTTAATTCAGG	GTGATTTCTCTGGAGGTTGA	ATAATTTTGGAAATATCAGAC	1080	
Db	1021	CATAAACTGTTAATTCAGG	GTGATTTCTCTGGAGGTTGA	ATAATTTTGGAAATATCAGAC	1080	
QY	1081	ACAGCTGATTAACAGGGA	GTGAAGAGGCTGCAATGAC	CACTTCTATTAGTTTGC	1140	
Db	1081	ACAGCTGATTAACAGGGA	GTGAAGAGGCTGCAATGAC	CACTTCTATTAGTTTGC	1140	
QY	1141	GAGGCATTTGATAAACT	GAATCCTTGTCTCTGGAA	TTATAGATCAGCTGAGTGTGAT	1200	
Db	1141	GAGGCATTTGATAAACT	GAATCCTTGTCTCTGGAA	TTATAGATCAGCTGAGTGTGAT	1200	
QY	1201	CCCAATAGTAAATGAAC	CTCTTAAAGTAACTGCAAG	TGTGTACATACAGACATGACGA	1260	
Db	1201	CCCAATAGTAAATGAAC	CTCTTAAAGTAACTGCAAG	TGTGTACATACAGACATGACGA	1260	
QY	1261	CTTGTGTTGGTCTGTAAG	ATGAGTATGTTTGTACCT	GCACACAGACGGCCATA	1320	
Db	1261	CTTGTGTTGGTCTGTAAG	ATGAGTATGTTTGTACCT	GCACACAGACGGCCATA	1320	
QY	1321	GTACAGCTGTGTCAGAG	GGGAAACACATGCTCAG	AAGGTTGGCCACCTCACAGAACACTC	1380	
Db	1321	GTACAGCTGTGTCAGAG	GGGAAACACATGCTCAG	AAGGTTGGCCACCTCACAGAACACTC	1380	
QY	1381	CGTGGTCATCGGAA	CAAAAGTCAATGTTGCT	TATATCTCTCATCAGGCTCTCAGCTCGGTAT	1440	
Db	1381	CGTGGTCATCGGAA	CAAAAGTCAATGTTGCT	TATATCTCTCATCAGGCTCTCAGCTCGGTAT	1440	
QY	1441	GATCAAGATATCTGAT	TATCTCGAGGTGGAATTTT	CAGTCAATTTTGGGACATATTT	1500	
Db	1441	GATCAAGATATCTGAT	TATCTCGAGGTGGAATTTT	CAGTCAATTTTGGGACATATTT	1500	
QY	1501	TCCTGAGAAATGAAAC	ATATCTTCTGTGTTCA	TGTGTGAGATTACTCAACTCTTAGTT	1560	
Db	1501	TCCTGAGAAATGAAAC	ATATCTTCTGTGTTCA	TGTGTGAGATTACTCAACTCTTAGTT	1560	
QY	1561	CCACCTGAAATCTGTAG	TGCAAGATGACATCTG	CTCTGTGTAGCCAGTGAACAC	1620	
Db	1561	CCACCTGAAATCTGTAG	TGCAAGATGACATCTG	CTCTGTGTAGCCAGTGAACAC	1620	
QY	1621	TCAGTAGGACTTCTAAG	TTTGGAGAAATGCAAT	ATGTTGGCATCTCGTCACTT	1680	
Db	1621	TCAGTAGGACTTCTAAG	TTTGGAGAAATGCAAT	ATGTTGGCATCTCGTCACTT	1680	
QY	1681	TTTCTCTATTCAAG	TAAATGAGGCTTCTG	ATGATTAATCTGGTGGGGTGTCA	1740	
Db	1681	TTTCTCTATTCAAG	TAAATGAGGCTTCTG	ATGATTAATCTGGTGGGGTGTCA	1740	
QY	1741	GATGTTCTGTGTAG	CTGTGGCAAAATGGAT	ATCTGTGTGTGTGTGTGGG	1800	
Db	1741	GATGTTCTGTGTAG	CTGTGGCAAAATGGAT	ATCTGTGTGTGTGTGTGGG	1800	
QY	1801	ATAACAGAGTGTGAG	ATTTAAACGCTTGTG	ATGAAGCTGCTGCTGTGATTTCA	1860	
Db	1801	ATAACAGAGTGTGAG	ATTTAAACGCTTGTG	ATGAAGCTGCTGCTGTGATTTCA	1860	
QY	1861	CTTAGTCTCCAGCAG	CTCAACCTTAAACAAAG	CTATGACGAGACGTAGTCTTGTGCTCTT	1920	
Db	1861	CTTAGTCTCCAGCAG	CTCAACCTTAAACAAAG	CTATGACGAGACGTAGTCTTGTGCTCTT	1920	
QY	1921	AAAAATATGCGCCAT	CATTAAGCTAATAA	CCCTTGGCAACTAA	CCTCTTGGCTTCTGAGGCA	1980
Db	1921	AAAAATATGCGCCAT	CATTAAGCTAATAA	CCCTTGGCAACTAA	CCTCTTGGCTTCTGAGGCA	1980
QY	1981	TCCTGACAGGGGAA	TTTACATAATTTCT	CACTCCCTGATGTT	CAAGCAATAAAG	2040
Db	1981	TCCTGACAGGGGAA	TTTACATAATTTCT	CACTCCCTGATGTT	CAAGCAATAAAG	2040
QY	2041	ACAAACTTAACAG	ACCCGGACATACAT	GTGCTATTCTTTGAT	TGGAAGCTTGTATT	2100
Db	2041	ACAAACTTAACAG	ACCCGGACATACAT	GTGCTATTCTTTGAT	TGGAAGCTTGTATT	2100
QY	2101	CAACTCCTGACT	GAAGAGCTCTAG	CGGCAATACTGCT	TATTTTCCCGAGAGATTTG	2160
Db	2101	CAACTCCTGACT	GAAGAGCTCTAG	CGGCAATACTGCT	TATTTTCCCGAGAGATTTG	2160
QY	2161	CAAAAAGCATCTGG	CAGTTTACAAAGGGG	CTCTTTTAACTG	CGAAAAACGAGCAGCA	2220
Db	2161	CAAAAAGCATCTGG	CAGTTTACAAAGGGG	CTCTTTTAACTG	CGAAAAACGAGCAGCA	2220
QY	2221	GTTCTCTTCCAA	CNAGTGAAAGAAC	CGATCAAAAGAG	AACATCAAGGAAACACCTCTTGTAT	2280

Qy 4441 CATGCGGAAGGAGCAGCGCTTCATGGCTAA 4473
 }
 Db 4441 CATGCGGAAGGAGCAGCGCTTCATGGCTAA 4473

RESULT 2

US-10-093-463-153
; Sequence 153, Application US/10093463
; Publication No. US20030208039A1

: GENERAL INFORMATION:

APPLICANT: Padigar, Muralidhara	60/274, 194
APPLICANT: Shenoy, Suresh	60/274, 194
APPLICANT: Kekuda, Ramesh	60/274, 194
APPLICANT: Gusev, Vladimir	60/274, 194
APPLICANT: Pochart, Pascal	60/274, 194
APPLICANT: Zhong, Mei	60/274, 194
APPLICANT: Rastelli, Luca	60/274, 194
APPLICANT: Mezes, Peter	60/274, 194
APPLICANT: Smithson, Glennda	60/274, 194
APPLICANT: Guo, Xiaojia	60/274, 194
APPLICANT: Casman, Valerie	60/274, 194
APPLICANT: Geisach, Stacie	60/274, 194
APPLICANT: Boldog, Ferenc	60/274, 194
APPLICANT: Li, Li	60/274, 194
APPLICANT: Zerhusen, Bryan	60/274, 194
APPLICANT: Tchernev, Velizar	60/274, 194
APPLICANT: Gangolli, Esha	60/274, 194
APPLICANT: Vernet, Corine	60/274, 194
APPLICANT: Pena, Carol	60/274, 194
APPLICANT: Burgess, Catherine	60/274, 194
APPLICANT: Liu, Xiaohong	60/274, 194
APPLICANT: Spytek, Kimberly	60/274, 194
APPLICANT: Gorman, Linda	60/274, 194
APPLICANT: Spaderna, Steven	60/274, 194
APPLICANT: Voss, Edward	60/274, 194
APPLICANT: Malyankar, Uriel	60/274, 194
APPLICANT: Anderson, David	60/274, 194
APPLICANT: Patturajan, Meera	60/274, 194
APPLICANT: Miller, Charles	60/274, 194
APPLICANT: Taupier, Raymond J. Jr.	60/274, 194
TITLE OF INVENTION: No. US2003020803	60/274, 194
TITLE OF INVENTION: Encoding the An	60/274, 194
FILE REFERENCE: 21402-290A (Cura 590	60/274, 194
CURRENT APPLICATION NUMBER: US/10/09	60/274, 194
CURRENT FILING DATE: 2002-06-24	60/274, 194
PRIOR APPLICATION NUMBER: 60/283, 675	60/274, 194
PRIOR FILING DATE: 2001-04-14	60/274, 194
PRIOR APPLICATION NUMBER: 60/338, 092	60/274, 194
PRIOR FILING DATE: 2001-12-03	60/274, 194
PRIOR APPLICATION NUMBER: 60/274, 281	60/274, 194
PRIOR FILING DATE: 2001-03-08	60/274, 194
PRIOR APPLICATION NUMBER: 60/274, 101	60/274, 194
PRIOR FILING DATE: 2001-03-08	60/274, 194
PRIOR APPLICATION NUMBER: 60/325, 681	60/274, 194
PRIOR FILING DATE: 2001-09-27	60/274, 194
PRIOR APPLICATION NUMBER: 60/304, 354	60/274, 194
PRIOR FILING DATE: 2001-07-10	60/274, 194
PRIOR APPLICATION NUMBER: 60/279, 995	60/274, 194
PRIOR FILING DATE: 2001-03-30	60/274, 194
PRIOR APPLICATION NUMBER: 60/294, 899	60/274, 194
PRIOR FILING DATE: 2001-05-31	60/274, 194
PRIOR APPLICATION NUMBER: 60/287, 424	60/274, 194
PRIOR FILING DATE: 2001-04-30	60/274, 194
PRIOR APPLICATION NUMBER: 60/299, 027	60/274, 194
PRIOR FILING DATE: 2001-06-18	60/274, 194
PRIOR APPLICATION NUMBER: 60/309, 198	60/274, 194
PRIOR FILING DATE: 2001-07-31	60/274, 194
PRIOR APPLICATION NUMBER: 60/281, 194	60/274, 194
PRIOR FILING DATE: 2001-04-04	60/274, 194
PRIOR APPLICATION NUMBER: 60/274, 194	60/274, 194
PRIOR FILING DATE: 2001-03-08	60/274, 194
PRIOR APPLICATION NUMBER: 60/274, 849	60/274, 194

Db 740 TCCGATGCCGAGACTATCTCTGTGTGTTTCAGGTCTTAGTGAATGACAGCATGG 799
Qy 781 ACCGGGGGGGACTTTGTCTCATCAGATAAAGTCAATTTGGACAGAAAATGGGCAAGT 840
Db 800 ACCGGGGGGGACTTTGTCTCATCAGATAAAGTCAATTTGGACAGAAAATGGGCAAGT 859
Qy 841 TATATTTACAACTACCTGCCAGTTGGCTTCAGCTAGTGATTCATTTCCGCGAGTGATGG 900
Db 860 TATATTTACAACTACCTGCCAGTTGGCTTCAGCTAGTGATTCATTTCCGCGAGTGATGG 919
Qy 901 GGGAAAGCAGTTGAAATTTAAATTCCTCTGTACAAATATCTCTCTGGATCGAAAAGAT 960
Db 920 GGGAAAGCAGTTGAAATTTAAATTCCTCTGTACAAATATCTCTCTGGATCGAAAAGAT 979
Qy 961 AAAGAGTTGCTAAATTTGCTCTCTGTTACTCGGTTCTTCTATGGATCGAGAAATATTTTC 1020
Db 980 AAAGAGTTGCTAAATTTGCTCTCTGTTACTCGGTTCTTCTATGGATCGAGAAATATTTTC 1039
Qy 1021 CATAACTGTTAAATTCAGGTGATCTCTCTGAAAGTTGAATATTTGGAAATATTCAGAC 1080
Db 1040 CATAACTGTTAAATTCAGGTGATCTCTCTGAAAGTTGAATATTTGGAAATATTCAGAC 1099
Qy 1081 ACAGCTGATAAAGGGAAGTGAAGAGGCTGCAATGACATCTTCTATTTAGTTGCAA 1140
Db 1100 ACAGCTGATAAAGGGAAGTGAAGAGGCTGCAATGACATCTTCTATTTAGTTGCAA 1159
Qy 1141 GAGGCATTTGATAAATCTGATCTCTGCTGGAATATAGATCAGCTGAGTGTGATTT 1200
Db 1160 GAGGCATTTGATAAATCTGATCTCTGCTGGAATATAGATCAGCTGAGTGTGATTT 1219
Qy 1201 CCCAATAGTAAATGAACTCTTAAAGTAACTGCAAGTGTGATACATACAGACATGGAACGA 1260
Db 1220 CCCAATAGTAAATGAACTCTTAAAGTAACTGCAAGTGTGATACATACAGACATGGAACGA 1279
Qy 1261 CTTGTTTGTGCTGAGATGAGNAGCATAGTTATTTGTAATCTGCAACAGAGCGGCATTA 1320
Db 1280 CTTGTTTGTGCTGAGATGAGNAGCATAGTTATTTGTAATCTGCAACAGAGCGGCATTA 1339
Qy 1321 GTACAGCTGTTGCAAGGGGAACATCTCTCAGAAAGGTTGGCCACTCACAGAACTC 1380
Db 1340 GTACAGCTGTTGCAAGGGGAACATCTCTCAGAAAGGTTGGCCACTCACAGAACTC 1399
Qy 1381 CGTGGTCAATCGGAACAAAGTCAATGTTGCTATATCTCATCAGGTCTCAGCTCGGTAT 1440
Db 1400 CGTGGTCAATCGGAACAAAGTCAATGTTGCTATATCTCATCAGGTCTCAGCTCGGTAT 1459
Qy 1441 GATCAAGATACCTGATATCTGAGGTGTGGATTTTTCAGTCAATATTTGGGACATATTT 1500
Db 1460 GATCAAGATACCTGATATCTGAGGTGTGGATTTTTCAGTCAATATTTGGGACATATTT 1519
Qy 1501 TCTGGGAAATGAAACATATCTCTGTGTTTCATGGTGTGAGATTAATCACTCTAGTT 1560
Db 1520 TCTGGGAAATGAAACATATCTCTGTGTTTCATGGTGTGAGATTAATCACTCTAGTT 1579
Qy 1561 CCACCTGAAATCTGATGCAAGAGTACAGCACTGCAATCTGCTGTAGCCAGTACCCAC 1620
Db 1580 CCACCTGAAATCTGATGCAAGAGTACAGCACTGCAATCTGCTGTAGCCAGTACCCAC 1639
Qy 1621 TCAGTAGGACTTCTAAGTTTGGAGAGAAATGCAATATTTGGGACATCTCTGCACTTT 1680
Db 1640 TCAGTAGGACTTCTAAGTTTGGAGAGAAATGCAATATTTGGGACATCTCTGCACTTT 1699
Qy 1681 TTTCTTATTCAGTAAATCAATGAGGCTTCTGATGATTAATCTGTTGGTGGGCTTCA 1740
Db 1700 TTTCTTATTCAGTAAATCAATGAGGCTTCTGATGATTAATCTGTTGGTGGGCTTCA 1759
Qy 1741 GATGGTTCTGTGACGTCGCAAAATGATCTGTTGATTTGGATCTGTTGATGGGG 1800
Db 1760 GATGGTTCTGTGACGTCGCAAAATGATCTGTTGATTTGGATCTGTTGATGGGG 1819
Qy 1801 ATAAACAGCAGTTGAGATTTAAACGCTTGTGATGAAGCTGTTCTCTGCTGTTGATTA 1860

Db 1820 ATAAACAGCAGTTGAGATTTAAACGCTTGTGATGAAGCTGTTCTCTGCTGCTGTTGATTA 1879
Qy 1861 CTTTAGTCAATCCAGCAGTCAACCTAAACCAAGCTATGACGAGACGTAGTCTTGCTCTTT 1920
Db 1880 CTTTAGTCAATCCAGCAGTCAACCTAAACCAAGCTATGACGAGACGTAGTCTTGCTCTTT 1939
Qy 1921 AAAAATATGGCCCATCATTAAGCTACAAAACCTTTGCAACTAAACCTCTTGCTGCTGAGGCA 1980
Db 1940 AAAAATATGGCCCATCATTAAGCTACAAAACCTTTGCAACTAAACCTCTTGCTGCTGAGGCA 1999
Qy 1981 TCTGCAAGGGAATTTTACCTAAATATTTCTATAACTCCCTGATGTTTCAAGCAATAAAG 2040
Db 2000 TCTGCAAGGGAATTTTACCTAAATATTTCTATAACTCCCTGATGTTTCAAGCAATAAAG 2059
Qy 2041 ACAAACTTAACAGACCCGGAACATACATGCTGCTATTTCTTGTGATGGAAGCGTTGATTA 2100
Db 2060 ACAAACTTAACAGACCCGGAACATACATGCTGCTATTTCTTGTGATGGAAGCGTTGATTA 2119
Qy 2101 CAACTCTGACTGAAAGCGCTCTAGCCGGAATACCTGCTTATTTTCTCCAGAGAAATTTG 2160
Db 2120 CAACTCTGACTGAAAGCGCTCTAGCCGGAATACCTGCTTATTTTCTCCAGAGAAATTTG 2179
Qy 2161 CAAAAAGCATCTGGCAGTTTCAGACAAAGGGGGCTCTTTTAACTGGAACCGAGCAGCA 2220
Db 2180 CAAAAAGCATCTGGCAGTTTCAGACAAAGGGGGCTCTTTTAACTGGAACCGAGCAGCA 2239
Qy 2221 GTTCTCTTCCAAACAGTGAAGAAACGATCAAGAGAAACATCAAGGAAACACCTCTTGTAT 2280
Db 2240 GTTCTCTTCCAAACAGTGAAGAAACGATCAAGAGAAACATCAAGGAAACACCTCTTGTAT 2299
Qy 2281 GATGAAGAGAGATGAGAGATATGAGCGAGAGAGGGAAGAAAGTGAATCTGTAATAT 2340
Db 2300 GATGAAGAGAGATGAGAGATATGAGCGAGAGAGGGAAGAAAGTGAATCTGTAATAT 2359
Qy 2341 CGGTCCAGCAAAATCAAGCCATTTGACCTATTTAGATATAATTTTAATGGAACCTGCA 2400
Db 2360 CGGTCCAGCAAAATCAAGCCATTTGACCTATTTAGATATAATTTTAATGGAACCTGCA 2419
Qy 2401 AAGCTGTTTATGCTCTGCTCTCAACGCTGGGGTTTGAATGAAGTACTGGAATGAAGTTTC 2460
Db 2420 AAGCTGTTTATGCTCTGCTCTCAACGCTGGGGTTTGAATGAAGTACTGGAATGAAGTTTC 2479
Qy 2461 TTGGATCGCTTTGGAATGCTGAAACCCCACTGCAACGATTCGTTTGGCTCTTGTCAAGA 2520
Db 2480 TTGGATCGCTTTGGAATGCTGAAACCCCACTGCAACGATTCGTTTGGCTCTTGTCAAGA 2539
Qy 2521 GGAGGCCATATGCTCACTGATGCTGCGGGTTAATCAAGCTGCTTTGTAACCTGCACAT 2580
Db 2540 GGAGGCCATATGCTCACTGATGCTGCGGGTTAATCAAGCTGCTTTGTAACCTGCACAT 2599
Qy 2581 GGGMAAACAGAAAGTAGAAGGAAGCTGCCAGCGCTCTGAGGAGTAGAAGGGAACCTTAC 2640
Db 2600 GGGMAAACAGAAAGTAGAAGGAAGCTGCCAGCGCTCTGAGGAGTAGAAGGGAACCTTAC 2659
Qy 2641 GGAGTGTCCCGTCCCGTCAACACAGCATCTCTGTCTATCATTTCTTTGGCAAAATACT 2700
Db 2660 GGAGTGTCCCGTCCCGTCAACACAGCATCTCTGTCTATCATTTCTTTGGCAAAATACT 2719
Qy 2701 TTAATGAGTATGACCAATGCAACTTTTATTTGTTGATCATATGAAGAGGCTCTACAGG 2760
Db 2720 TTAATGAGTATGACCAATGCAACTTTTATTTGTTGATCATATGAAGAGGCTCTACAGG 2779
Qy 2761 CCACCTTAGAACCAAGCAACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAACTCCAGT 2820
Db 2780 CCACCTTAGAACCAAGCAACCCAGACCTTTCTAAGGCAAGGGGTTCCCTCCAACTCCAGT 2839
Qy 2821 AATATTTGCAAGGACAGATTAATAAAGTTGCTGCACTGCTGTTTCCGCTCGGCTGAT 2880
Db 2840 AATATTTGCAAGGACAGATTAATAAAGTTGCTGCACTGCTGTTTCCGCTCGGCTGAT 2899
Qy 2881 GCTGATCACTCTGGCTCTGACCCCTCTCTGCTCTGCTTTACATCACTGTTTCTTAGTA 2940
Db 2900 GCTGATCACTCTGGCTCTGACCCCTCTCTGCTCTGCTTTACATCACTGTTTCTTAGTA 2959

```
QY 2941 AATGAAGTTGGAGTCAGTTAGCTGCTATGCACTGTGTTATGCTGCGACACCTACTGCGA 3000
Db 2960 AATGAAGTTGGAGTCAGTTAGCTGCTATGCACTGTGTTATGCTGCGACACCTACTGCGA 3019
QY 3001 TTGGATAAAATTTAGGGCTCCCTTTCTGGAGATGCTGGCCGGAAGATGGCAAGATCGATGC 3060
Db 3020 TTGGATAAAATTTAGGGCTCCCTTTCTGGAGATGCTGGCCGGAAGATGGCAAGATCGATGC 3079
QY 3061 TTGGAGGTGAGAGAGCGGACAGAGCGCTGCTTCTGGCGGAATCTGAGAGAAATTTAGCAG 3120
Db 3080 TTGGAGGTGAGAGAGCGGACAGAGCGCTGCTTCTGGCGGAATCTGAGAGAAATTTAGCAG 3139
QY 3121 GCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180
Db 3140 GCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3199
QY 3181 GTCATATCACTCGAGTCAATCAGAGAGCGCGCAGACTATCAACACGGCTCCTGATGCC 3240
Db 3200 GTCATATCACTCGAGTCAATCAGAGAGCGCGCAGACTATCAACACGGCTCCTGATGCC 3259
QY 3241 TCAGGGCTGAAGCAAAAGTCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
Db 3260 TCAGGGCTGAAGCAAAAGTCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3319
QY 3301 ACTGGTTGCTTATCAAGTGTCCACAAATGAAAATAATTTCTACATCTTACGAGGAAGA 3360
Db 3320 ACTGGTTGCTTATCAAGTGTCCACAAATGAAAATAATTTCTACATCTTACGAGGAAGA 3379
QY 3361 CGAAGCAAGCTACCGCTATTTGTTTACTTGGAGTAATAGAGCTGAAATTTGCTGTGAA 3420
Db 3380 CGAAGCAAGCTACCGCTATTTGTTTACTTGGAGTAATAGAGCTGAAATTTGCTGTGAA 3439
QY 3421 ATTGAACCTCCTAAACTAATTTGACAGACCTCGAAGCTCTAGCCAAATTTCTGAGGATTC 3480
Db 3440 ATTGAACCTCCTAAACTAATTTGACAGACCTCGAAGCTCTAGCCAAATTTCTGAGGATTC 3499
QY 3481 GGGTTGACTAGTGGTGGATCCAACTACTCGCTGGCCAGACATCTTGCAGGCACTGACG 3540
Db 3500 GGGTTGACTAGTGGTGGATCCAACTACTCGCTGGCCAGACATCTTGCAGGCACTGACG 3559
QY 3541 TTTCTTCTGCTACAGCCTCCAAAGCCCAAACTTCTCTCCACACAGACATCTCCGAAAGACA 3600
Db 3560 TTTCTTCTGCTACAGCCTCCAAAGCCCAAACTTCTCTCCACACAGACATCTCCGAAAGACA 3619
QY 3601 GCCATTGATCTGATTTGACGCTGGGTTCACTGTTTGGAGCCTTACATGATGTGCTCGCT 3660
Db 3620 GCCATTGATCTGATTTGACGCTGGGTTCACTGTTTGGAGCCTTACATGATGTGCTCGCT 3679
QY 3661 GTTCTGATGGGCTTCTCGAACTTTGTCGATGCGCGAGAAACAACTTGCACACATCACA 3720
Db 3680 GTTCTGATGGGCTTCTCGAACTTTGTCGATGCGCGAGAAACAACTTGCACACATCACA 3739
QY 3721 ATGGGGTTGCTCTGAGCCAGACGCTGACTCGGCGGCTCTGCGAGGATGCCCTCTCG 3780
Db 3740 ATGGGGTTGCTCTGAGCCAGACGCTGACTCGGCGGCTCTGCGAGGATGCCCTCTCG 3799
QY 3781 CTCATTGCCACCGCAGACACCGCTTCTCATCCACCATAGCCAAAGAGGTACACAGA 3840
Db 3800 CTCATTGCCACCGCAGACACCGCTTCTCATCCACCATAGCCAAAGAGGTACACAGA 3859
QY 3841 CATACGGCTTTGCGAGCAATACCCATCTCAGGAGATGTCACACAACTCTTGCA 3900
Db 3860 CATACGGCTTTGCGAGCAATACCCATCTCAGGAGATGTCACACAACTCTTGCA 3919
QY 3901 CGAGCTAAAGGGGAAATTTTGGAGTCAATTTGAAATTTCTTAAAGAGATGCCACAGAT 3960
Db 3920 CGAGCTAAAGGGGAAATTTTGGAGTCAATTTGAAATTTCTTAAAGAGATGCCACAGAT 3979
QY 3961 GTTGTGGATCTTCTCGTGGAGGTTATGGAATCATCATATGATGCTGCTTGAAGGATCTTTA 4020
Db 3980 GTTGTGGATCTTCTCGTGGAGGTTATGGAATCATCATATGATGCTGCTTGAAGGATCTTTA 4039
```

```
QY 4021 GTTAAAAAGAAAGGTCTTCAAGAAATGTTTCCAGCAATCTGCAAGTTCTTACATGTCAGC 4080
Db 4040 GTTAAAAAGAAAGGTCTTCAAGAAATGTTTCCAGCAATCTGCAAGTTCTTACATGTCAGC 4099
QY 4081 TATTATGAGCGGAATACAGAAATAGCAATAGCAATGAGCTCGCATGTTCAAGTGGCCCTGTAC 4140
Db 4100 TATTATGAGCGGAATACAGAAATAGCAATGAGCTCGCATGTTCAAGTGGCCCTGTAC 4159
QY 4141 GATATCCGACCTGGAAATGTGAGCAATCTCATGAGCAACAGGGACCAATCACTGCAGTG 4200
Db 4160 GATATCCGACCTGGAAATGTGAGCAATCTCATGAGCAACAGGGACCAATCACTGCAGTG 4219
QY 4201 GCTTTTGTCTCTGATGGAAGATATCTTGCCACTACTCAAACTGACAGCCACATTTCT 4260
Db 4220 GCTTTTGTCTCTGATGGAAGATATCTTGCCACTACTCAAACTGACAGCCACATTTCT 4279
QY 4261 TTTTGGCAGATGAACACGTCACCTGCTGGGAAGCATCGGATGCTGAACTCGGCACCTCAG 4320
Db 4280 TTTTGGCAGATGAACACGTCACCTGCTGGGAAGCATCGGATGCTGAACTCGGCACCTCAG 4339
QY 4321 CTGGCTGCTATTAACCTTACCAGGTGCCCTGTGAGCCGCGTCCCGGCTCCAC 4380
Db 4340 CTGGCTGCTATTAACCTTACCAGGTGCCCTGTGAGCCGCGTCCCGGCTCCAC 4399
QY 4381 AATGCCCTCAAGCTGCGCGGCTCATCTGAGCTTCCAAACCGCAACGTCTCATCTGCGC 4440
Db 4400 AATGCCCTCAAGCTGCGCGGCTCATCTGAGCTTCCAAACCGCAACGTCTCATCTGCGC 4459
QY 4441 CATGACGGGAAGAGACCGCTTCTCATGTCCTAA 4473
Db 4460 CATGACGGGAAGAGACCGCTTCTCATGTCCTAA 4492
```

RESULT 3

```
US-10-120-988-91
; Sequence 91, Application US/10120988
; Publication No. US20030219745A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Ren, Feiyun
; APPLICANT: Wang, Dunrui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No US20030219745A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 802CON
; CURRENT APPLICATION NUMBER: US/10/120,988
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 09/774,528
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 91
; LENGTH: 6224
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3597)
US-10-120-988-91
```

```
Query Match 78.2%; Score 3497; DB 17; Length 6224;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3508; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 966 GTTGCTAAATTTGTCCTCTGTTACTCGGTTCTTATGATGATGAGAGATATTTCCATAA 1025
Db 117 GTTGCTAAATTTGTCCTCTGTTACTCGGTTCTTCTATGATGATGAGAGATATTTCCATAA 176
QY 1026 ACTGTTAAATTCAGGGTGATTTCTTCGAAAGGTTGAATATTTGGAACATATCAGACAGC 1085
Db 177 ACTGTTAAATTCAGGGTGATTTCTTCGAAAGGTTGAATATTTGGAACATATCAGACAGC 236
```


Db 2397 GCTGAGCAAAAGTCCAGAGGAGAGCATGACCTGTTGTGACGATGATCACCACCTGG 2456
Qy 3306 TTGCTTATCAAGTGTCCCAAAATGAAAAAATTTCTCATCTTACGAGAAAGACGAA 3365
Db 2457 TTGCTTATCAAGTGTCCCAAAATGAAAAAATTTCTCATCTTACGAGAAAGACGAA 2516
Qy 3366 GCAAGCTACCGCTATTGTTTACTTGGAGTATAGGAGCTGAATTTGGTCTGGAATTGA 3425
Db 2517 GCAAGCTACCGCTATTGTTTACTTGGAGTATAGGAGCTGAATTTGGTCTGGAATTGA 2576
Qy 3426 ACCTCTTAAACTATTGACAGAGCTCGAAGCTCTAGCCAAATTCCTGAGGGAATTCGGGTT 3485
Db 2577 ACCTCTTAAACTATTGACAGAGCTCGAAGCTCTAGCCAAATTCCTGAGGGAATTCGGGTT 2636
Qy 3486 GACTAGTGGTGGATCAACTTACTCGCTGGCCAGACATATTGCAAGGCACTGACGTTTCT 3545
Db 2637 GACTAGTGGTGGATCAACTTACTCGCTGGCCAGACATATTGCAAGGCACTGACGTTTCT 2696
Qy 3546 TCTGCTACAGCTCCGAAGCCCAAACTTCTCCACACAGCACTATCCGAAGACAGCCAT 3605
Db 2697 TCTGCTACAGCTCCGAAGCCCAAACTTCTCCACACAGCACTATCCGAAGACAGCCAT 2756
Qy 3606 TGATCTGATTGGAGCTGGGTTCACTGTTGGGAGCCTTACATGGAGTGTCCGCTGTTCT 3665
Db 2757 TGATCTGATTGGAGCTGGGTTCACTGTTGGGAGCCTTACATGGAGTGTCCGCTGTTCT 2816
Qy 3666 GATGGGGCTTCTCGAACTTTGTGGCCGATCCGAGAAACAACTTGGCCAAATCACAATGGG 3725
Db 2817 GATGGGGCTTCTCGAACTTTGTGCCGATCCGAGAAACAACTTGGCCAAATCACAATGGG 2876
Qy 3726 GTTGCCCTGAGCCGACGAGCTGACTCGGCCGCTCTGGAGGCACTGCCCTCTCGCTCAT 3785
Db 2877 GTTGCCCTGAGCCGACGAGCTGACTCGGCCGCTCTGGAGGCACTGCCCTCTCGCTCAT 2936
Qy 3786 TGCACCGCCAGACACCGCCCTTCACTACACCATAGCAAGAGGTACACAGACATAC 3845
Db 2937 TGCACCGCCAGACACCGCCCTTCACTACACCATAGCAAGAGGTACACAGACATAC 2996
Qy 3846 GGTCTTTGACGAAATACCAATACAGAGAAATATGACACAACTTGGCAGCAGC 3905
Db 2997 GGTCTTTGACGAAATACCAATACAGAGAAATATGACACAACTTGGCAGCAGC 3056
Qy 3906 TAAAGGGGAATTTTTCAGAGTCAATGAAATTTCTTATTGAAAGATCCCAAGATGCTCT 3965
Db 3057 TAAAGGGGAATTTTTCAGAGTCAATGAAATTTCTTATTGAAAGATCCCAAGATGCTCT 3116
Qy 3966 GGATCTTCTGAGGTTATGACATCAATATGATCTGCTTGAAGGATCTTTAGTTAA 4025
Db 3117 GGATCTTCTGAGGTTATGACATCAATATGATCTGCTTGAAGGATCTTTAGTTAA 3176
Qy 4026 AAGAAAGTCTTCAAGAAATTTTCCAGCAATCTGAGGTTCTAATGTCAGCTATTAA 4085
Db 3177 AAGAAAGTCTTCAAGAAATTTTCCAGCAATCTGAGGTTCTAATGTCAGCTATTAA 3236
Qy 4086 TGAAGGGAATACAGAAATAGCAGTGGAGTCCGATGTTTCAAGTCCGCTGTCAGACAT 4145
Db 3237 TGAAGGGAATACAGAAATAGCAGTGGAGTCCGATGTTTCAAGTCCGCTGTCAGACAT 3296
Qy 4146 CCGGATCGGAAATGTGAGCAATCCATGACACAAAGGAGCAATCACTGCACTGCTTTT 4205
Db 3297 CCGGATCGGAAATGTGAGCAATCCATGACACAAAGGAGCAATCACTGCACTGCTTTT 3356
Qy 4206 TGCTCTGTATGGAAGATATCTTGGCACTTCACTCAAACTGACAGCAATTTCTTTTGG 4265
Db 3357 TGCTCTGTATGGAAGATATCTTGGCACTTCACTCAAACTGACAGCAATTTCTTTTGG 3416
Qy 4266 GCAAGTGAACAGCTCTGCTGGAGAGCATCGGCATGCTGAATCCGCACTCAGCTGG 4325
Db 3417 GCAAGTGAACAGCTCTGCTGGAGAGCATCGGCATGCTGAATCCGCACTCAGCTGG 3476
Qy 4326 CTGCATTTAAACCTACAGGTGCCCTCTGTGAGCCCGGCTCCCGGGTCCCAATGC 4385

Db 3477 CTGCATTTAAACCTACAGGTGCCCTCTGTGAGCCCGGCTCCCGGGTCCCAATGC 3536
Qy 4386 CTTCAAGCTTGGCCGGCTCATCTGGACTTCCAAACGCAACGCTCATCTCATGGCCCATG 4444
Db 3537 CTTCAAGCTTGGCCGGCTCATCTGGACTTCCAAACGCAACGCTCATCTCATGGCCCATG 3596
Qy 4445 ACGGGAAGGAGCACCCTTTCATGGTCTAA 4473
Db 3597 ACGGGAAGGAGCACCCTTTCATGGTCTAA 3625

RESULT 4
US-10-094-749-1317
; Sequence 1317, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOMYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1317
; LENGTH: 3316
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-1317

Query Match 62.8%; Score 2807.4; DB 17; Length 3316;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 3080; Conservative 0; Mismatches 1; Indels 242; Gaps 3;
Qy 1 ATGGCAGGAAACAGCCTTGTCTACCCATTGTTCTTTGGGGTCGAAAAGCGCCACACAT 60
Db 236 ATGGCAGGAAACAGCCTTGTCTACCCATTGTTCTTTGGGGTCGAAAAGCGCCACACAT 295
Qy 61 TGCATCTCAGCGGTACTTTTAAACAGATATGGGGCCACGATCGTAAACAGATGTACGAC 120
Db 296 TGCATCTCAGCGGTACTTTTAAACAGATATGGGGCCACGATCGTAAACAGATGTACGAC 355
Qy 121 GCACAAATATGTCCTGGGATCTTTTCAGTAGAACCTGCAATTAATCCTCAGCACCTGTTG 180
Db 356 GCACAAATATGTCCTGGGATCTTTTCAGTAGAACCTGCAATTAATCCTCAGCACCTGTTG 415
Qy 181 TTGGTTCATACAGCATCAATCACTTGTGTTGTCTAAAGCTTGTGCTTCCAGTGCAAAACAG 240
Db 416 TTGGTTCATACAGCATCAATCACTTGTGTTGTCTAAAGCTTGTGCTTCCAGTGCAAAACAG 475
Qy 241 TATATTGTGAGTGCATCTGAAAGTGGAGATGTGCTCTGGGATGTGAGTGTGAGTGCAGA 300
Db 476 TATATTGTGAGTGCATCTGAAAGTGGAGATGTGCTCTGGGATGTGAGTGTGAGTGCAGA 535

QY 301 TGTATTGAATTTACAAAATTAGCTTGACACATATCTGGCATACAGTTCTACCGAGTTCTCT 360
DB 536 TGTATTGAATTTACAAAATTAGCTTGACACATATCTGGCATACAGTTCTACCGAGTTCTCT 595
QY 361 GTTGGGAATCAGCGAGAAAGGCTTTTATGCCACGGACATTACCTGGAATCCCTGT 420
DB 596 GTTGGGAATCAGCGAGAAAGGCTTTTATGCCACGGACATTACCTGGAATCCCTGT 655
QY 421 GTGGATGCTACCGCTTGAAGTATTATATCTCTCTTAGTATCAAGATATACCCAGACTGG 480
DB 656 GTGGATGCTACCGCTTGAAGTATTATATCTCTCTTAGTATCAAGATATACCCAGACTGG 715
QY 481 ATTAGCTCCATGAGTATTATTTCGATCCACCGAACACAGAGACACAGTGGTAGCACTC 540
DB 716 ATTAGCTCCATGAGTATTATTTCGATCCACCGAACACAGAGACACAGTGGTAGCACTC 775
QY 541 TCGGTGACTGGCATCTGGAAGGCTGAGATTGTTACCTCGGAAATTAAGTGACATGCAAGGAT 600
DB 776 TCGGTGACTGGCATCTGGAAGGCTGAGATTGTTACCTCGGAAATTAAGTGACATGCAAGGAT 835
QY 601 ACTGAGCCAAATTTGAGGAGGAATCCAAACCAATTTATTGTGAGAAATTCGCAAGCATC 660
DB 836 ACTGAGCCAAATTTGAGGAGGAATCCAAACCAATTTATTGTGAGAAATTCGCAAGCATC 895
QY 661 TCTTTTGTGCAATTTACAAAGGTCACTTTTGGTGTGTGTTCCAAATATTGAGGGTG 720
DB 896 TCTTTTGTGCAATTTACAAAGGTCACTTTTGGTGTGTGTTCCAAATATTGAGGGTG 955
QY 721 TTGATGCGCGAGACTTCTCTGTTGTTGTTGAGTCTGAGTCCAGTGAATGAGACAGATGG 780
DB 956 TTGATGCGCGAGACTTCTCTGTTGTTGTTGAGTCTGAGTCCAGTGAATGAGACAGATGG 1015
QY 781 ACCGGGGGAGCTTTGTCTCATCAGATAAAGTCAATTTGAGACAGAAATGGCAAGT 840
DB 1016 ACCGGGGGAGCTTTGTCTCATCAGATAAAGTCAATTTGAGACAGAAATGGCAAGT 1075
QY 841 TATATTACAACTACTGCGAGTTCCTGTTCCAGCTAGTGATTCATTTCCGCACTGATGG 900
DB 1076 TATATTACAACTACTGCGAGTTCCTGTTCCAGCTAGTGATTCATTTCCGCACTGATGG 1135
QY 901 GGGAGGCGAGTTGAAATTTAAATCTCTGTTCAACATATCTCTGATCCAAAGAT 960
DB 1136 GGGAGGCGAGTTGAAATTTAAATCTCTGTTCAACATATCTCTGATCCAAAGAT 1195
QY 961 AAAGAGTTGCTAATTTGTCTCTCTGTTTACTCGGTCTTCTATGATGACAGAAATTTTC 1020
DB 1196 AAAGAGTTGCTAATTTGTCTCTCTGTTTACTCGG-TCTTCTATGATGACAGAAATTTTC 1254
QY 1021 CATAAACTGTTAATCAGGTTGATTTCTTGGAGGTTGAATTTTGGAACTATCAGAC 1080
DB 1255 CATAAACTGTTAATCAGGTTGATTTCTTGGAGGTTGAATTTTGGAACTATCAGAC 1314
QY 1081 ACAGCTGATAAACAGGAGTGAAGAGGCTGGCAATGCAACTTCTATTAGTTGGAA 1140
DB 1315 ACAGCTGATAAACAGGAGTGAAGAGGCTGGCAATGCAACTTCTATTAGTTGGAA 1374
QY 1141 GAGGCAATTTGATAAACTGAACTCTGTTCTGCTGGAATTTATAGATCAGCTGAGTGTGATT 1200
DB 1375 GAGGCAATTTGATAAACTGAACTCTGTTCTGCTGGAATTTATAGATCAGCTGAGTGTGATT 1434
QY 1201 CCCAATAGTAATGAACCTCTTAAAGTAACTGCAAGTGTGTATACACAGACATGGACGA 1260
DB 1435 CCCAATAGTAATGAACCTCTTAAAGTAACTGCAAGTGTGTATACACAGACATGGACGA 1494
QY 1261 CTTGTTTGGTCTGTAAGTGAAGCATAGTTATTGTACTCTGCCACAGCGGCATTA 1320
DB 1495 CTTGTTTGGTCTGTAAGTGAAGCATAGTTATTGTACTCTGCCACAGCGGCATTA 1554
QY 1321 GTACAGCTGTTGACGGGGAACATGCTCAGAAAGGTTGGCCACTCAGCAACATC 1380
DB 1555 GTACAGCTGTTGACGGGGAACATGCTCAGAAAGGTTGGCCACTCAGCAACATC 1614

QY 1381 CGTGTCTATCGGAACAAAGTCAATGTTTGTCTATATCTCTCATCAGGTCTCAGCTCGGTAT 1440
DB 1615 CGTGTCTATCGGAACAAAGTCAATGTTTGTCTATATCTCTCATCAGGTCTCAGCTCG --- 1670
QY 1441 GATCAAAGATACCTGATATCTCGAGGTGTGGAATTTTTCAGTCATATAATTTGGGACATATTT 1500
DB 1671 ----- 1670
QY 1501 TCTGGAGAAATGAAACATATCTTCTGTGTTTCATGGTGTGAGATTACTCAACTTCTAGTT 1560
DB 1671 ----- 1670
QY 1561 CCACCTGAAAACTGTAGTGCAGAGTACAGCACTGCACTGTCTCTGTAGCCAGTGACCAC 1620
DB 1671 -----GCAAGAGTACAGCACTGCACTGTCTCTGTAGCCAGTGACCAC 1712
QY 1621 TCAGTAGGACTTCTAAAGTTTGCAGAGAAAAAATGCAATAATGTTGGCATCTCGTCACTT 1680
DB 1713 TCAGTAGGACTTCTAAAGTTTGCAGAGAAAAAATGCAATAATGTTGGCATCTCGTCACTT 1772
QY 1681 TTTCTTATTCAAGTAATCAAAATGGAGCCCTTCTGTATGATTTACCTGGTGGGGTGTTC 1740
DB 1773 TTTCTTATTCAAGTAATCAAAATGGAGCCCTTCTGTATGATTTACCTGGTGGGGTGTTC 1832
QY 1741 GATGGTCTGTGTACGTCTGGCAAAATGGATACATGTTGTCATTTGGATCGTTGTGTGATGGG 1800
DB 1833 GATGGTCTGTGTACGTCTGGCAAAATGGATACATGTTGTCATTTGGATCGTTGTGTGATGGG 1892
QY 1801 ATAAACAGCAGTGTGAGATTCTAAACGCTTGTGATGAAGCTGTCTCTGCTGTGTGATTCA 1860
DB 1893 ATAAACAGCAGTGTGAGATTCTAAACGCTTGTGATGAAGCTGTCTCTGCTGTGTGATTCA 1952
QY 1861 CTTAGTCTATCCAGCAGTCAACCTTAAACAGCTATGACAGAGCTAGTCTTGTGCTCTT 1920
DB 1953 CTTAGTCTATCCAGCAGTCAACCTTAAACAGCTATGACAGAGCTAGTCTTGTGCTCTT 2012
QY 1921 AAAATATGGCCCATCATTAAGCTACAAACCTTTGCACTAACTCTTGGCTTCTGAGGCA 1980
DB 2013 AAAATATGGCCCATCATTAAGCTACAAACCTTTGCACTAACTCTTGGCTTCTGAGGCA 2072
QY 1981 TCTGCAAGGGAAATTTTACATAATTTCTATAACTCTCCCTGATGTTCAAGCAATAAAG 2040
DB 2073 TCTGCAAGGGAAATTTTACATAATTTCTATAACTCTCCCTGATGTTCAAGCAATAAAG 2132
QY 2041 ACAAACCTTAACAGACCCGACATACATGCTATTCTTGTGATGGAAGCGTTGATTAAT 2100
DB 2133 ACAAACCTTAACAGACCCGACATACATGCTATTCTTGTGATGGAAGCGTTGATTAAT 2192
QY 2101 CAACTCTGATGGAAGAGCCTCTAGGCGGAATACCTGCTTTATTTTCCCGAGAGAAATTTG 2160
DB 2193 CAACTCTGATGGAAGAGCCTCTAGGCGGAATACCTGCTTTATTTTCCCGAGAGAAATTTG 2252
QY 2161 CAAAAGCATCTGGCAGTTTCAGCAAGGGGGCTCTTTTAACTCGAAACAGCAGCA 2220
DB 2253 CAAAAGCATCTGGCAGTTTCAGCAAGGGGGCTCTTTTAACTCGAAACAGCAGCA 2312
QY 2221 GTTCTCTTCCAACTGGAAGAAACGATCAAGAGAAACATCAAGGAAACACCTCTTGTAT 2280
DB 2313 GTTCTCTTCCAACTGGAAGAAACGATCAAGAGAAACATCAAGGAAACACCTCTTGTAT 2372
QY 2281 GATGAAGAGGAGATGAGAGATTAATGAGGAGAGAGAGGAGAAAGTGAATCTGTAATAT 2340
DB 2373 GATGAAGAGGAGATGAGAGATTAATGAGGAGAGAGAGGAGAAAGTGAATCTGTAATAT 2432
QY 2341 CGGTCCAGCAATCAAGCCATTGACCTTATTAGATATAATTTTAACTATGGAACCTGCA 2400
DB 2433 CGGTCCAGCAATCAAGCCATTGACCTTATTAGATATAATTTTAACTATGGAACCTGCA 2492
QY 2401 AAGCTGTTTATGTTCTGCTCTTCAAGCTGGGGTTTGAATGAAGTACTGATGAAGTTTGC 2460
DB 2493 AAGCTGTTTATGTTCTGCTCTTCAAGCTGGGGTTTGAATGAAGTACTGATGAAGTTTGC 2552
QY 2461 CTGGATCGCCTTTGGAATGCTGAAACCCCACTGCACCGTATCGTTTGGCCTCTTGTCAAGA 2520

```
Db 2553 CTGGATCGCTTGGAAATGCTGAACCCCACTGACCGTATCGTTGGCCCTCTGTCAAGA 2612
QY 2521 GGAGGCCATATGTCATCTGATGTCGCGGGTTAATACAGCTGCTTTGTAACATGTCACAT 2580
Db 2613 GGAGGCCATATGTCATCTGATGTCGCGGGTTAATACAGCTGCTTTGTAACATGTCACAT 2672
QY 2581 GGGAAACAGAGTAGGAGGAGCTGCCAGCGTCTGAGGAGTAGGAAGGAACTTAC 2640
Db 2673 GGGAAACAGAGTAGGAAGGAGCTGCCAGCGTCTGAGGAGTAGGAAGGAACTTAC 2732
QY 2641 GGAAGTCTCCGTCGCTGCTACCAACACAGCATCTCTCTGTCATCATTTCTTTGGCAATACT 2700
Db 2733 GGAAGTCTCCGTCGCTGCTACCAACACAGCATCTCTCTGTCATCATTTCTTTGGCAATACT 2792
QY 2701 TTAATGAGTATGACCAATGCACTTTTATTTGGTGATCATATGAAGAAGGCTCTACAGG 2760
Db 2793 TTAATGAGTATGACCAATGCACTTTTATTTGGTGATCATATGAAGAAGGCTCTACAGG 2852
QY 2761 CCACCTAGACCAAGCACCAGACCTTTCTAAGGCAAGGGTTCCCTCCACTTCCAGT 2820
Db 2853 CCACCTAGACCAAGCACCAGACCTTTCTAAGGCAAGGGTTCCCTCCACTTCCAGT 2912
QY 2821 AATATTGTGCAAGACAGATTTAAACAAAGTTGCTGCACCTGTCGTTCCGCTCGGTCTGAT 2880
Db 2913 AATATTGTGCAAGACAGATTTAAAC----- 2937
QY 2881 GCTGATCACTCTGGCTCTGACCCCTCTCTGCTCTGCTTTTACATACCTGTTTCTAGTA 2940
Db 2938 ----- 2937
QY 2941 AATGAAGTTGGAGTCAGTTAGCTGCTATGCTATGCACTGTGTTATGCTGCGACCTACTGGGA 3000
Db 2938 ----AAGGTTGGAGTCAGTTAGCTGCTATGCTATGCACTGTGTTATGCTGCGACCTACTGGGA 2993
QY 3001 TTGGATAAATTTAGGCTCCCTTTCTGGAGATGCTGCGCGCAAGATGCGAAGATCGATGC 3060
Db 2994 TTGGATAAATTTAGGCTCCCTTTCTGGAGATGCTGCGCGCAAGATGCGAAGATCGATGC 3053
QY 3061 TTGGAGGTGAGAGAAGCCGACAGGCGCTGCTTTCTGCGGAACTGAGAAGAAATTGAGCAG 3120
Db 3054 TTGGAGGTGAGAGAAGCCGACAGGCGCTGCTTTCTGCGGAACTGAGAAGAAATTGAGCAG 3113
QY 3121 GCAGGCAGGAAGAGCCATTGATGCTGCGGCTCTTACTTACTTACTAGTACATAGACCCAC 3180
Db 3114 GCAGGCAGGAAGAGCCATTGATGCTGCGGCTCTTACTTACTTACTAGTACATAGACCCAC 3173
QY 3181 GTCATATCACTGGAGTCACATCAGAAAGCCGCGCAGACTATCACCGGCTCCTGATGCC 3240
Db 3174 GTCATATCACTGGAGTCACATCAGAAAGCCGCGCAGACTATCACCGGCTCCTGATGCC 3233
QY 3241 TCAGGGCCTGAAGCAAAAGTCCAGGAGGAAGAGATGACCTTTGTCAGATGACATCACC 3300
Db 3234 TCAGGGCCTGAAGCAAAAGTCCAGGAGGAAGAGATGACCTTTGTCAGATGACATCACC 3293
QY 3301 ACTGGTGTCTTATCAAGTGTCCTCC 3323
Db 3294 ACTGGTGTCTTATCAAGTGTCCTCC 3316
```

RESULT 5
US-10-108-260A-1051
; Sequence 1051, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cdna
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108, 260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1051

```
; LENGTH: 2578  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-108-260A-1051  
  
Query Match 10.6%; Score 474.6; DB 17; Length 2578;  
Best Local Similarity 55.1%; Pred. No. 5.1e-132;  
Matches 999; Conservative 0; Mismatches 799; Indels 15; Gaps 3;  
  
QY 29 TTGTTCTTTGGGGTGGAAAGCGCCACACATTCGATCTCAGGGTACTTTTAAACAGATG 88  
Db 276 TGGCACATCTCTGGGACAGAAAGGCCCTCCCCACAGCATCACTGCCATCATGATCATGTG 335  
QY 89 ATGGGSCCAGATCGTAAACAGAGATGTCAGACGGAATAATATGTCTCTGGGATCTTTTCA 148  
Db 336 ACAGGGAACGATTTGTGACTTGAAGTCAAGAGGGTCAGCTCTGTCTCTGGAATCTCTCAC 395  
QY 149 TAGAACTGCAAAATTAATCTCGAGCACTGTTGTTGGTCATACAGCATCAATCACTTGT 208  
Db 396 ATGAACTAAAGATTTTCAGCGAAAGAACTCTTATTTGGTCAATTCAGCTTCGGTAAACATG 455  
QY 209 TGTCTAAAGCTTGTGTTCCAGTGACAAACAGATATATTGTGAGTGCATCTGAAAGTGGAG 268  
Db 456 TGSCAAGAGCAAGGGACATTTCTCTAAACAGCCCTACATTTGTTAGTGTCTGTAATAATGG 515  
QY 269 AGATGTGCTCTGGGATGAGTGATGCGAGATGTAATTCGAAATTTACAAAATTAGCTTGCA 328  
Db 516 AGATGTGCTTTGGAAATGTCACCAATGGACAGTGCCTGGAGAGGCTACACTTCCTTACA 575  
QY 329 CACATACCTGGCATACAGTTCTACAGTTCTCTGTTGGGAATCAGCGAAGAGGAGCTTT 388  
Db 576 GGCACACTGCAATCTGTTATTACCACTGCTGCTTCGGATGACAGGAGAGGCTGGCTTC 635  
QY 389 TATGCCACGGAATTAACCTGAAATCTTGTGTGATGCTACCAGCCTTGAAGTATTAT 448  
Db 636 TTTGTTGTGGAGAATATCAAGATGTCCTTATAATTTGATGCCAAAACCTTTGGCTGTTG 695  
QY 449 ACTCTCTTAGTATCAAGATATCACAGACTGGAATAGCTCCATGATGATATTATTCGATCCC 508  
Db 696 ACAGTTTATGATCATCTCAGTTTCTGACTGGATCAACTGCAATGTCATGTTTCACTCCA 755  
QY 509 ACCGAACACAGAGGACACAGTGGTAGCACTCTCGGTGATCGGCATCTTGAAGGTCTGGA 568  
Db 756 TGAGAAATTCAGAAAGATTTCTCTCTGGTGTATCAGTAGCTGTGTAGCTCAAAAGTATGG 815  
QY 569 TTGTTACCTCGGAATAAGTGACATGACAGTACTGAGCCCAATATTTGAGAGGAGATCCA 628  
Db 816 ATCTTTCTCTCATCTATCAACAGCAATTCAGGAAAGCAAGATGTCATGAAAAAGAAATCCA 875  
QY 629 AACCAGTTTATGTCAGAAATTCGCAAGCATCTCTTTTGTGCAATTTACAAAGGTCCAC 688  
Db 876 AGTTTCTTGAGTCTTGAATCTGCCAGCAATTCGATTTTGCATATATCTGAGAGACTTC 935  
QY 689 TTTTGGTTCTGTGTTCCAAATATTTGAGGGTGTTCGATGCCGAGACTATTCCTGTTGT 748  
Db 936 TATTGGTGTGTTTCTTAAATGTTGGAAGTTTATGATTTATGATTTTTCCTCTCTGC 995  
QY 749 GTTCAGGTCTTAGTGAATAATGGACACATGAGCCGGGGGGACTTTGTCATCAGATA 808  
Db 996 TGACTGAAGTTAGTAGAAATGGGCAGTTCTTTTGTGAGAGAGTGAATGCTGCTCACA 1055  
QY 809 AGTGCATCATTTGGACAGAAAAATGGGCAAGTTATATTTTACAAACTACCTGCCAGTGGCC 868  
Db 1056 GAATCTCATCTGGACAGAAAGATGGTCAACATGTTACATCTATCAGCTGCTGAAACAGTGGC 1115  
QY 869 TTCCAGCTAGTGAATTCATTTCCGAGTGTGGGGAAGGAGTGTGAAAAATTTTAAATTCCTC 928  
Db 1116 TTTCAAAAAG-----CATATACCTGCTGATGGAAGAGTGTCTTAAAGAGACCATTTATC 1169  
QY 929 CTGTACAAATATCTCTTGGATCGAAAAAGATGAAGTGTGTTAAATTTGCTCTCTCTGTTA 988  
Db 1170 CTCATTTACTGTGCTCTACTTCTGTGCGAGAAAAATGAAGAACAGAGCCGCTCTTGTGTTA 1229
```

```
QY 989 CTGGTCTCTTATGGATGACAGAAATATTTCCATAAATGTTAAATTCAGGGTGATCTTT 1048
Db 1230 TGGCTACATGAATGAAGAAAGAGCCCTTTTACAAAGTACTTTTCTCTGGAGAGTCT 1289
QY 1049 CTGAAGGTTGAATATTTGGAACATATCAGACAGCTG-----ATAAACAGGGAAGTG 1102
Db 1290 CAGGAGAAATATCTTTGTGACATCCCTGATGTTCTGTATCCAGTTTGTATGTTCTC 1349
QY 1103 AAGAGGGCTGGCAATGACAACTCTTATATAGTTTGAAGGCAATTTGATAAATCAATC 1162
Db 1350 CTAGAGAGATACCAGTAAGTCCACCTGGACTCTTCAAGATAAATTTTGAAGCATGATA 1409
QY 1163 CTGTCTCTGCTGGAATATAGATCAGCTGAGTGTGATCCCATAGTAAATGAACCTTTA 1222
Db 1410 CTATGTCACAAAGTATTAATGACTATTCTCTGGCTTAAAGATGGGGCAGGAACGCTG 1469
QY 1223 AAGTAACCTGCAAGTGTGTACATACACAGCATGACGACTGTTTGTGGTCTGGAAGATG 1282
Db 1470 TAGTCACTTCATCAGAGTATATTTCCAGTCTTGATAAATTAATATGTTGGCTGTGAAGTG 1529
QY 1283 GAAGCATAGTATATGACCTGCCACACAGACGCGCATAGTACAGCTGTTGCAAGGGGAAC 1342
Db 1530 GGACAAATATATATCCAGGCTTGAATGCTGCCAAAGCAAGACTTCTGGAAGTGGTT 1589
QY 1343 ACATGCTCAGAGAGGTTGGCCACCTCCACAGACATCCGTGTCATCGGAACAAAGTCA 1402
Db 1590 CTTTAGT- --AAGAGATTTCCCTCCCTCATAAAGTCTTAAAGGCCACCAAGTGTCA 1646
QY 1403 CATGTTTGTCTATATCTCATCAGTCTCAGCTCGGTATGATCAAGATACCTCATATCTG 1462
Db 1647 CTTCAATTAATCTACCAACATGGTCTCTCTTGGAAATAGACAAAGTTGGATGTTCTG 1706
QY 1463 GAGGTGTGGAATTTTTCAGTCAATAATTTGGGACATATTTTCTGGAGAAATGAACAATATCT 1522
Db 1707 GGGACCTGGACTCATGTGTGATCTTGTGGATATCTTTACTGAAGAAATTTTGCATAAAT 1766
QY 1523 TCTGTGTTATGTTGGTGTGAGATTAATCACTTCTAGTTCACCTCGAACTGAAAGCTGAGTGCAA 1582
Db 1767 TCTTTTGGAGCTGGTCCAGTAACAAAGTCTTTTGTGATGTCAACAGAGAGTTTAAACTAA 1826
QY 1583 GAGTACAGCACTGCATCTGCTGTAGCCAGTACCACTCAGTAGGACTCTTAAGTTTGC 1642
Db 1827 GGGGTGAGCAGATAATTTGCTGTGTGCGGTGACCATTCGTTGGCTCTCTTCACTTG 1886
QY 1643 GAGAGAAAAATGCAATAATTTGGCATCTCGTCACCTTTTCTCTATTTCAAGTAATCAAT 1702
Db 1887 AGGGAAGAGTTGCTCTGATGCGCGGAGCACTTTTCTGTGAGGATGATAAAT 1946
QY 1703 GAGAGGCTTCTGATGATTAATCTGCTGGTGGGTTGTTCAAGTGTCTGTGAGTCTGGC 1762
Db 1947 GGCACCCGGTTGAGAAATTTTAAATTTGTTGGATGTGCAGATGACTCAGTTTATATCTGG 2006
QY 1763 AAATGATACCTGTGATGATGATGCTGTGTGATGGGATAACAGCAGTTGAGATCTAA 1822
Db 2007 AAATGAAACAGGCACTTTGGAAGACATGAGACAGAGAGAGACAGCAAGTAATTTCTTA 2066
QY 1823 ACGCTTGTGTGATGA 1835
Db 2067 ATTGTTGTGATGA 2079
```

RESULT 6

```
US-10-029-386-7862
; Sequence 7862, Application US/10029386
; Publication No. US200301947041
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: ABOICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
```

```
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7862
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR18.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: P22813, EVALUE 2.90e-01
; OTHER INFORMATION: NT HIT: AF188125.1, EVALUE 3.00e-64
; OTHER INFORMATION: EST_HUMAN HIT: AW896708.1, EVALUE 7.00e-20
US-10-029-386-7862
```

```
Query Match 3.4%; Score 150.2; DB 16; Length 585;
Best Local Similarity 74.9%; Pred. No. 6.3e-34;
Matches 188; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
```

```
QY 635 TTTATTGTGCAGAATTCGCCAAAGCATCTCTTTTGTGCAATTTACACAAAGGTCACTTTTGG 694
Db 186 TTTTATTTTAACTTGAATGTGCTGATCTTACTTTTACAAATATGGAAGTAGATT 245
QY 695 TTGTGTGTTCCAAATATTGGAGGGTGTTCGATGCCGGAGACTATTCTTGTGTGTTCCAG 754
Db 246 GTGAATTTCTTCTTCTTTCAGGTTGTCGATGCCGGAGACTATTCTTGTGTGTTCCAG 305
QY 755 GTCTAGTGAATAATGACAGACATGACCGGGGGGACTTTGTCTCATCAGATAAAGTCA 814
Db 306 GTCTAGTGAATAATGACAGACATGACCGGGGGGACTTTGTCTCATCAGATAAAGTCA 365
QY 815 TCATTTGGACAGAAATGGCAAAATGTTATTTTACAACTACTCTGCCAGTTGCTTCCAG 874
Db 366 TCATTTGGACAGAAATGGCAAAATGTTATTTTACAACTACTCTGCCAGTTATGCAGCAA 425
QY 875 CTAGTGATTTCA 885
Db 426 GTAAATAAATTA 436
```

RESULT 7

```
US-10-029-386-21562
; Sequence 21562, Application US/10029386
; Publication No. US200301947041
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: ABOICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21562
; LENGTH: 227
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR18.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: P22813, EVALUE 3.00e-01
; OTHER INFORMATION: NT HIT: AF188125.1, EVALUE 5.00e-64
; OTHER INFORMATION: EST_HUMAN HIT: AW896708.1, EVALUE 2.00e-20
US-10-029-386-21562
```



```
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-320926

Query Match      2.8%; Score 125.4; DB 13; Length 590;
Best Local Similarity 99.2%; Pred. No. 2.1e-26;
Matches 126; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 593 TGCAGGATCTGAGCCCAATTTGAGGAGGAATCCAAACCAATTTATGTCAGAAATGCC 652
    |||||
Db 346 TTCAGGATCTGAGCCCAATTTGAGGAGGAATCCAAACCAATTTATGTCAGAAATGCC 405

QY 653 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTGGTGTGTGTTCCAAATATT 712
    |||||
Db 406 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTGGTGTGTGTTCCAAATATT 465

QY 713 GGAGGGT 719
    |||||
Db 466 GGAGGGT 472

RESULT 11
US-10-027-632-320926
; Sequence 320926, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastsEQ for Windows Version 4.0
; SEQ ID NO 320926
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-320926

Query Match      2.8%; Score 125.4; DB 17; Length 590;
Best Local Similarity 99.2%; Pred. No. 2.1e-26;
Matches 126; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 593 TGCAGGATCTGAGCCCAATTTGAGGAGGAATCCAAACCAATTTATGTCAGAAATGCC 652
    |||||
Db 346 TTCAGGATCTGAGCCCAATTTGAGGAGGAATCCAAACCAATTTATGTCAGAAATGCC 405

QY 653 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTGGTGTGTGTTCCAAATATT 712
    |||||
Db 406 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTGGTGTGTGTTCCAAATATT 465

QY 713 GGAGGGT 719
    |||||
Db 466 GGAGGGT 472

RESULT 12
US-10-027-632-237407
; Sequence 237407, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
```

```
; Sequence 237407, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastsEQ for Windows Version 4.0
; SEQ ID NO 237407
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-237407

Query Match      2.8%; Score 125.4; DB 13; Length 598;
Best Local Similarity 99.2%; Pred. No. 2.2e-26;
Matches 126; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 593 TGCAGGATCTGAGCCCAATTTGAGGAGGAATCCAAACCAATTTATGTCAGAAATGCC 652
    |||||
Db 241 TTCAGGATCTGAGCCCAATTTGAGGAGGAATCCAAACCAATTTATGTCAGAAATGCC 300

QY 653 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTGGTGTGTGTTCCAAATATT 712
    |||||
Db 301 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTGGTGTGTGTTCCAAATATT 360

QY 713 GGAGGGT 719
    |||||
Db 361 GGAGGGT 367

RESULT 13
US-10-027-632-237408
; Sequence 237408, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
```



```
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 237408
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-237408

Query Match      2.8%; Score 125.4; DB 13; Length 598;
Best Local Similarity 99.2%; Pred. No. 2.2e-26;
Matches 126; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 593 TGCAGGATACGTAGCCCAATATTGAGGAGGAATCCAAACCAATTTATTGTCAGAAATTGCC 652
Db 241 TTCCAGGATACGTAGCCCAATATTGAGGAGGAATCCAAACCAATTTATTGTCAGAAATTGCC 300

QY 653 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 712
Db 301 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 360

QY 713 GGAGGGT 719
Db 361 GGAGGGT 367

RESULT 14
US-10-027-632-237407
; Sequence 237407, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 237407
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-237407

Query Match      2.8%; Score 125.4; DB 17; Length 598;
Best Local Similarity 99.2%; Pred. No. 2.2e-26;
Matches 126; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 593 TGCAGGATACGTAGCCCAATATTGAGGAGGAATCCAAACCAATTTATTGTCAGAAATTGCC 652
Db 241 TTCCAGGATACGTAGCCCAATATTGAGGAGGAATCCAAACCAATTTATTGTCAGAAATTGCC 300

QY 653 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 712
Db 301 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 360

QY 713 GGAGGGT 719
Db 361 GGAGGGT 367

RESULT 15
US-10-027-632-237408
; Sequence 237408, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 237408
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-237408

Query Match      2.8%; Score 125.4; DB 17; Length 598;
Best Local Similarity 99.2%; Pred. No. 2.2e-26;
Matches 126; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 593 TGCAGGATACGTAGCCCAATATTGAGGAGGAATCCAAACCAATTTATTGTCAGAAATTGCC 652
Db 241 TTCCAGGATACGTAGCCCAATATTGAGGAGGAATCCAAACCAATTTATTGTCAGAAATTGCC 300

QY 653 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 712
Db 301 AAAGCATCTCTTTTGTGCATTTACACAAAGGTCACCTTTTGGTGTGTGTTCCAAATATT 360

QY 713 GGAGGGT 719
Db 361 GGAGGGT 367

Search completed: July 2, 2005, 13:41:56
Job time : 2395 secs
```


Qy	1146	ATTGTGATAAACTGAATCCTTTGTCCTGCTGGAAATTATAGATCAGCTGAGTGTGATTCCCAA	1205
Db	297	ATTGTGATAAACTGAATCCTTTGTCCTGCTGGAAATTATAGATCAGCTGAGTGTGATTCCCAA	356
Qy	1206	TAGTAATGAACCTCTTAAAGTAACCTGCAAGTGTGTATACACAGCATGAGACGACTTGT	1265
Db	357	TAGTAATGAACCTCTTAAAGTAACCTGCAAGTGTGTATACACAGCATGAGACGACTTGT	416
Qy	1266	TTGTGTCGTGAGATCGAAGCATAGTTATTTGTACCTGCCACACAGACGCGCCATAGTACA	1325
Db	417	TTGTGTCGTGAGATCGAAGCATAGTTATTTGTACCTGCCACACAGACGCGCCATAGTACA	476
Qy	1326	GCTGTTCCAAAGGGGAAACATGCTCAGAAGAGTTGGCCACCTCACAAGAACACTCCGTGG	1385
Db	477	GCTGTTCCAAAGGGGAAACATGCTCAGAAGAGTTGGCCACCTCACAAGAACACTCCGTGG	536
Qy	1386	TCATCGGAAACAAAGTGCATGTTTGTCTATATCCTCATCAGGTCCTAGCTCGGTATGATCA	1445
Db	537	TCATCGGAAACAAAGTGCATGTTTGTCTATATCCTCATCAGGTCCTAGCTCGGTATGATCA	596
Qy	1446	AAGNATCCTGATATCTGGAGGTGAGATTTTTTCAGTCATAATTTTGGGACATATTTCTCG	1505
Db	597	AAGNATCCTGATATCTGGAGGTGAGATTTTTTCAGTCATAATTTTGGGACATATTTCTCG	656
Qy	1506	AGAAATCAAAACATATCTTCTGTGTTCACTGCTGTGAGATTACTCAACTCTAGTCCACC	1565
Db	657	AGAAATCAAAACATATCTTCTGTGTTCACTGCTGTGAGATTACTCAACTCTAGTCCACC	716
Qy	1566	TGAAAACTGTAGTGCAAAGAGTACAGCACTGCATCTGCTGTGTACCGACAGTGCACCTCAGT	1625
Db	717	TGAAAACTGTAGTGCAAAGAGTACAGCACTGCATCTGCTGTGTACCGACAGTGCACCTCAGT	776
Qy	1626	AGGACTTCTAAGTTTGGAGAGAAAAAATGCGATAATGTTGGCATCTGTCACACTTTTCC	1685
Db	777	AGGACTTCTAAGTTTGGAGAGAAAAAATGCGATAATGTTGGCATCTGTCACACTTTTCC	836
Qy	1686	TATTCAGTAATCAAAATGGAGGCTTCTGATGATTAACCTGCTGCTGGGTGTTTCAGATGG	1745
Db	837	TATTCAGTAATCAAAATGGAGGCTTCTGATGATTAACCTGCTGCTGGGTGTTTCAGATGG	896
Qy	1746	TTCTGTGTACGTCCTGGCAAAATGGATCTGGTGCATTTGGATCGTTGTGTGATGGGGATAAC	1805
Db	897	TTCTGTGTACGTCCTGGCAAAATGGATCTGGTGCATTTGGATCGTTGTGTGATGGGGATAAC	956
Qy	1806	AGCATTTGAGATTTCTAAACGCTTGTGATGAAGCTGTTCTGCTGCTGTTGATTCACCTTAG	1865
Db	957	AGCATTTGAGATTTCTAAACGCTTGTGATGAAGCTGTTCTGCTGCTGTTGATTCACCTTAG	1016
Qy	1866	TCATCCAGCAGTCAACCTTAAACCAAGCTATGACGAGACGTAGTCTTGCTGCTCTTAAAAA	1925
Db	1017	TCATCCAGCAGTCAACCTTAAACCAAGCTATGACGAGACGTAGTCTTGCTGCTCTTAAAAA	1076
Qy	1926	TATGGCCCATCATAAAGCTACAAACCCCTTGCAACTAACTCTTGCTTCTGAGGCATCTGA	1985
Db	1077	TATGGCCCATCATAAAGCTACAAACCCCTTGCAACTAACTCTTGCTTCTGAGGCATCTGA	1136
Qy	1986	CAAGGGAAATTTACTTAAATATTTCTCATATACTCCCTGATGGTTCAGCAATAAAGACAAA	2045
Db	1137	CAAGGGAAATTTACTTAAATATTTCTCATATACTCCCTGATGGTTCAGCAATAAAGACAAA	1196
Qy	2046	CCTTAACAGCCCGACATACATGCTGCTATTTCTTTGTATGGAGCGTTGATTAATCAACT	2105
Db	1197	CCTTAACAGCCCGACATACATGCTGCTATTTCTTTGTATGGAGCGTTGATTAATCAACT	1256
Qy	2106	CCTGACTGAAGAAGCCTCTAGGCCGAATATGCTCTTTATTTCCCGACAGAAATTTGCAAAA	2165
Db	1257	CCTGACTGAAGAAGCCTCTAGGCCGAATATGCTCTTTATTTCCCGACAGAAATTTGCAAAA	1316
Qy	2166	AGCATCTGGCAGTTTCAGACAAAGGGGCTCTTTTTTTAACTTGAAAAACGACGACGATTTCT	2225
Db	1317	AGCATCTGGCAGTTTCAGACAAAGGGGCTCTTTTTTTAACTTGAAAAACGACGACGATTTCT	1376

2226	Qy	CTTCCAAAGTGAAGAAACGATCAAGAGAAACATCAAGGAAACACTCTCTTGATGATGA	2288
1377	Db	CTTCCAAAGTGAAGAAACGATCAAGAGAAACATCAAGGAAACACTCTCTTGATGATGA	1436
2286	Qy	AGAGGAGGATGAGGAGATATGAGGCAGAGAGAGGGAAGAAAGTGATCCTGGAATATCGGTC	2345
1437	Db	AGAGGAGGATGAGGAGATATGAGGCAGAGAGAGGGAAGAAAGTGATCCTGGAATATCGGTC	1496
2346	Qy	CAGCAATCAAAAGCCATTGACCCCTATTAGAAATATAATTTAACTATATGGAACACTGCAAGCT	2405
1497	Db	CAGCAATCAAAAGCCATTGACCCCTATTAGAAATATAATTTAACTATATGGAACACTGCAAGCT	1556
2406	Qy	GTTTATGTCTGCTCTCAAGCCTTGGGGTTTGAATGAAGATGCTGGATGAAGTTTGGCTTGGGA	2465
1557	Db	GTTTATGTCTGCTCTCAAGCCTTGGGGTTTGAATGAAGATGCTGGATGAAGTTTGGCTTGGGA	1616
2466	Qy	TCGCCCTTGGAAATGCTGAAACCCCACTGCACCGTATCGTTTGGGCTCTTGTTCAGAGGAGG	2525
1617	Db	TCGCCCTTGGAAATGCTGAAACCCCACTGCACCGTATCGTTTGGGCTCTTGTTCAGAGGAGG	1676
2526	Qy	CCATATGTCACTGATGCTGCGGGTTATAATCAGCCTGCTGTGTAAACTGTGCACATGGGAA	2585
1677	Db	CCATATGTCACTGATGCTGCGGGTTATAATCAGCCTGCTGTGTAAACTGTGCACATGGGAA	1736
2586	Qy	AAACAGAAAGTAGGAAGAAAGTCCACCGTCTGAGGAGTAGGAAAGGGAACCTTACCGAGT	2645
1737	Db	AAACAGAAAGTAGGAAGAAAGTCCACCGTCTGAGGAGTAGGAAAGGGAACCTTACCGAGT	1796
2646	Qy	GTCCCGTGGCGTCAACACACACAGCATCTCTGCTCTATCATTTCTTTGGCAAAATACTTTAAT	2705
1797	Db	GTCCCGTGGCGTCAACACACACAGCATCTCTGCTCTATCATTTCTTTGGCAAAATACTTTAAT	1856
2706	Qy	GAGTATGACCAATGCAACTTTTATGCTGATCATATGAAGAAGGGTCTTACCAGGCGCAC	2765
1857	Db	GAGTATGACCAATGCAACTTTTATGCTGATCATATGAAGAAGGGTCTTACCAGGCGCAC	1916
2766	Qy	TAGACCAAGCACCCACAGACCTTTCTAAGGCAAGGGGTTCCTCCCTCCAACCTTCCAGTAATAT	2825
1917	Db	TAGACCAAGCACCCACAGACCTTTCTAAGGCAAGGGGTTCCTCCCTCCAACCTTCCAGTAATAT	1976
2826	Qy	TGTGCAAGGACAGATTAAACAAGTTGCTGCACTGTGCTGTTTCCGCTCGGTCTGATGCTGA	2885
1977	Db	TGTGCAAGGACAGATTAAACAAGTTGCTGCACTGTGCTGTTTCCGCTCGGTCTGATGCTGA	2036
2886	Qy	TCACTCTGGCTCGACCTCTTCTGCTCTGCTGTTTACATACCTGTTCTTAGTAATAATGA	2945
2037	Db	TCACTCTGGCTCGACCTCTTCTGCTCTGCTGTTTACATACCTGTTCTTAGTAATAATGA	2096
2946	Qy	AGGTTGGAGTCAGTTAGCTGTATGCACTGTGTTATGCTGCCAGACCTACTGGGATGGA	3005
2097	Db	AGGTTGGAGTCAGTTAGCTGTATGCACTGTGTTATGCTGCCAGACCTACTGGGATGGA	2156
3006	Qy	TAAATTTAGGCTTCCCTTCTGAGATGCTGCGCCGGAAGATGCAAGATCGATGCTTGA	3065
2157	Db	TAAATTTAGGCTTCCCTTCTGAGATGCTGCGCCGGAAGATGCAAGATCGATGCTTGA	2216
3066	Qy	GGTGAAGAGACCGCACAGGCCCTGCTTCTGGCGGAACCTGAGAAGAAATTCAGACAGCGCAGG	3125
2217	Db	GGTGAAGAGACCGCACAGGCCCTGCTTCTGGCGGAACCTGAGAAGAAATTCAGACAGCGCAGG	2276
3126	Qy	CAGGAAGGAAGCCATTGATGCCCTGGGCTCCTTACTTCACTCAGTACATAGACCAACGTCAT	3185
2277	Db	CAGGAAGGAAGCCATTGATGCCCTGGGCTCCTTACTTCACTCAGTACATAGACCAACGTCAT	2336
3186	Qy	ATCACTGGAGTCAATTCAGAAAGCGCGGACAGATATCAACAACGGCTCTGATGCTCAGG	3245
2337	Db	ATCACTGGAGTCAATTCAGAAAGCGCGGACAGATATCAACAACGGCTCTGATGCTCAGG	2396
3246	Qy	GCCTGAAGCAAAAGTCCAGAGGAGAGCATGACTTGTGTGACGATGACATCAACCTCGG	3305
2397	Db	GCCTGAAGCAAAAGTCCAGAGGAGAGCATGACTTGTGTGACGATGACATCAACCTCGG	2456
3306	Qy	TTGCTTATCAAGTGTCCCAAAAATGAAAAAATTTTCTTACATCTTACAGAGAAACGGA	3365

Db 2457 TTGCTTATCAAGTGTCCCAAAATGAAAAATTTCTACATCTTTACGAGGAAGACGGAA 2516
QY 3366 GCAAGCTACCGCTATTGTTTTTACTTGGAGTAATAGGAGCTGAAATTTGGTCTGCTGAAATTTGA 3425
Db 2517 GCAAGCTACCGCTATTGTTTTTACTTGGAGTAATAGGAGCTGAAATTTGGTCTGCTGAAATTTGA 2576
QY 3426 ACCTCTTAACCTATTGACGAGACTCGAAGCTCTAGCAAAATTCCTGAGGGATTCGGGTT 3485
Db 2577 ACCTCTTAACCTATTGACGAGACTCGAAGCTCTAGCAAAATTCCTGAGGGATTCGGGTT 2636
QY 3486 GACTAGTGGTGGATCCAACTACTCGCTGGCCAGACATACTTGCAGGCACTGACGTTTCT 3545
Db 2637 GACTAGTGGTGGATCCAACTACTCGCTGGCCAGACATACTTGCAGGCACTGACGTTTCT 2696
QY 3546 TCTGCTACAGCTCCAAAGCCCAAACTTCTCCACAGCAGCACTATCCGAGAAACAGCCAT 3605
Db 2697 TCTGCTACAGCTCCAAAGCCCAAACTTCTCCACAGCAGCACTATCCGAGAAACAGCCAT 2756
QY 3606 TGATCTGATGGAGTGGGTTCACTGTTGGGAGCCTTACATGATGTGTCGCTGTCT 3665
Db 2757 TGATCTGATGGAGTGGGTTCACTGTTGGGAGCCTTACATGATGTGTCGCTGTCT 2816
QY 3666 GATGGGCTTCTCGAACTTTGCGGATGCGGAGAAACAACTTGCACACATCAATGGG 3725
Db 2817 GATGGGCTTCTCGAACTTTGCGGATGCGGAGAAACAACTTGCACACATCAATGGG 2876
QY 3726 GTTGCTCTGAGCCAGCAGCTGACTCGGCCCGCTCTGCGAGGCATGCCCTCTCGCTCAT 3785
Db 2877 GTTGCTCTGAGCCAGCAGCTGACTCGGCCCGCTCTGCGAGGCATGCCCTCTCGCTCAT 2936
QY 3786 TGCACCGCCAGACACCGCCCTTCACTACCAATAGCAGCAATAGCAGGAGTACAGACATAC 3845
Db 2937 TGCACCGCCAGACACCGCCCTTCACTACCAATAGCAGCAATAGCAGGAGTACAGACATAC 2996
QY 3846 GGCCTTGCAGCAATACCAATCAGCAGCAATATGCACAACTCTTGCAGGAC 3905
Db 2997 GGCCTTGCAGCAATACCAATCAGCAGCAATATGCACAACTCTTGCAGGAC 3056
QY 3906 TAAAGGGGAAATTTGAGAGTCAATGAAATTTCTATTGAAAGATGCCACAGATGTTGT 3965
Db 3057 TAAAGGGGAAATTTGAGAGTCAATGAAATTTCTATTGAAAGATGCCACAGATGTTGT 3116
QY 3966 GGATCTTCTGAGGAGTTATGACATCATATTGTTACTGCTTGAAGGATCTTTAGTTAA 4025
Db 3117 GGATCTTCTGAGGAGTTATGACATCATATTGTTACTGCTTGAAGGATCTTTAGTTAA 3176
QY 4026 AAGAAAGTCTTCAAGATGTTTCCAGGCATCTGAGGTTCTACATGTCAGCTATTA 4085
Db 3177 AAGAAAGTCTTCAAGATGTTTCCAGGCATCTGAGGTTCTACATGTCAGCTATTA 3236
QY 4086 TGAGCGGAATCAGAAATAGCAGTGGAGCTCGCCATGTTTCACTGAGTGGCCCTGTACACAT 4145
Db 3237 TGAGCGGAATCAGAAATAGCAGTGGAGCTCGCCATGTTTCACTGAGTGGCCCTGTACACAT 3296
QY 4146 CCGAGCTGGAATGTGAGCAATCCATGAGCAAAAGGGACCAATCACTGCACTGGCTTT 4205
Db 3297 CCGAGCTGGAATGTGAGCAATCCATGAGCAAAAGGGACCAATCACTGCACTGGCTTT 3356
QY 4206 TGCTCTGATGGAAGATATCTGCCACCTACTCAAACTGACAGGCACATTTCTTTTG 4265
Db 3357 TGCTCTGATGGAAGATATCTGCCACCTACTCAAACTGACAGGCACATTTCTTTTG 3416
QY 4266 GCAGATGAACAGCTCACTGCTGGGAAGCATCGGCATGCTGAACTCGGCACCTCAGCTGG 4325
Db 3417 GCAGATGAACAGCTCACTGCTGGGAAGCATCGGCATGCTGAACTCGGCACCTCAGCTGG 3476
QY 4326 CTGCAATTAACCTACAGGTGCCCCCTGTGACGCGCGGCTCCCGGGTCCCAATGC 4385
Db 3477 CTGCAATTAACCTACAGGTGCCCCCTGTGACGCGCGGCTCCCGGGTCCCAATGC 3536
QY 4386 CTTCAAGCT-GGCCCGGCTCATCTGACTTCCAAACCGCAACGTCATCTCATGGCCCATG 4444

Db 3537 CCTCAAGCTGGGCGCGCTCATCTGCACTTCCAAACGCAACGTCATCTCTCATGGCCCATG 3596
QY 4445 ACGGGAAGGAGCACCAGCTTTTCATGGTCTAA 4473
Db 3597 ACGGGAAGGAGCACCAGCTTTTCATGGTCTAA 3625
RESULT 2
US-09-513-999C-14504
; Sequence 14504, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14504
; LENGTH: 356
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 13
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 20
; OTHER INFORMATION: k=g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 230
; OTHER INFORMATION: s=g or c
US-09-513-999C-14504
Query Match 2.3%; Score 103; DB 4; Length 356;
Best Local Similarity 100.0%; Pred. No. 1.3e-20;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCAGGAAACAGCCTTCTTACCCATTGTTCTTTGGGTGCGAAAGCGCCACACAT 60
Db 254 ATGGCAGGAAACAGCCTTCTTACCCATTGTTCTTTGGGTGCGAAAGCGCCACACAT 313
QY 61 TGCATCTCAGCGTACTTTTAAACAGATGATGGGCCACCATCG 103
Db 314 TGCATCTCAGCGTACTTTTAAACAGATGATGGGCCACCATCG 356
RESULT 3
US-08-188-582-17
; Sequence 17, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; NUCLEIC ACIDS ENCODING TAPs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2152 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2112
US-08-188-582-17

Query Match 1.1%; Score 48; DB 1; Length 2152;
Best Local Similarity 57.2%; Pred. No. 0.0031;
Matches 127; Conservative 0; Mismatches 90; Indels 5; Gaps 2;

QY 4034 GTCTTCAAGATGTTTCCAGCCATCTGCAGGTTCTACATGGTCAGGTTATATGACGGGA 4093
DB 1566 GCCTTTAAGATATTTGCCGGCCATCTTGCTGATGTAATTTGACCAGATTCCATCCAA 1625
QY 4094 ATCAGAGATAGCAGTTGGAGCTCGCC--ATGGTTTCAGTGGCCCTGTACGACATCCGGAC 4151
DB 1626 TTCTAATTATGTTGCTACGGGCTCTGCAGACAGACTGTCGGCTCTGGAGCTCTGAA 1685
QY 4152 TGGAAATGT---CAGACAAATCCATGACACAGGACCAATCACTGAGTGGCTTTTC 4208
DB 1686 TGGTAATGTTGAAGGATCTTCACTGGACACAGGACCAATTCATTTCTTGACATTTTC 1745
QY 4209 TCTGTATGGAAGATATCTTGGCCACCTACTCAAAACACTGACAG 4250
DB 1746 TCCCAATGGAGATCTCTGGCTACAGGACCAACAGATGGCAG 1787

RESULT 4
US-08-646-715-17
Sequence 17, Application US/08646715
Patent No. 5637686
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2152 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2112
US-08-646-715-17

Query Match 1.1%; Score 48; DB 1; Length 2152;
Best Local Similarity 57.2%; Pred. No. 0.0031;
Matches 127; Conservative 0; Mismatches 90; Indels 5; Gaps 2;

QY 4034 GTCTTCAAGATGTTTCCAGCCATCTGCAGGTTCTACATGGTCAGGTTATATGACGGGA 4093
DB 1566 GCCTTTAAGATATTTGCCGGCCATCTTGCTGATGTAATTTGACCAGATTCCATCCAA 1625
QY 4094 ATCAGAGATAGCAGTTGGAGCTCGCC--ATGGTTTCAGTGGCCCTGTACGACATCCGGAC 4151
DB 1626 TTCTAATTATGTTGCTACGGGCTCTGCAGACAGACTGTCGGCTCTGGAGCTCTGAA 1685
QY 4152 TGGAAATGT---CAGACAAATCCATGACACAGGACCAATCACTGAGTGGCTTTTC 4208
DB 1686 TGGTAATGTTGAAGGATCTTCACTGGACACAGGACCAATTCATTTCTTGACATTTTC 1745
QY 4209 TCTGTATGGAAGATATCTTGGCCACCTACTCAAAACACTGACAG 4250
DB 1746 TCCCAATGGAGATCTCTGGCTACAGGACCAACAGATGGCAG 1787

RESULT 5
US-09-806-708B-22
Sequence 22, Application US/09806708B
Patent No. 6784342
GENERAL INFORMATION:
APPLICANT: The University of British Columbia
TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
FILE REFERENCE: 4810-58741
CURRENT APPLICATION NUMBER: US/09/806,708B
CURRENT FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: US 60/147,133
PRIOR FILING DATE: 1999-08-04
NUMBER OF SEQ ID NOS: 23
SOFTWARE: Patent in version 3.0
SEQ ID NO 22


```
QY 2333 CTGAATATCGTCCAGCAAAATCAAGCCATTGACCCCTATTAGAAATAATTTAACTATCG 2392
Db 1266 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1207
QY 2393 ACACTCAAAAG 2403
Db 1206 RRRRRRRRRR 1196

RESULT 7
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. PAE1 promoters
US-09-806-708B-22

Query Match 0.9%; Score 39.2; DB 4; Length 1141;
Best Local Similarity 9.1%; Pred. No. 0.96;
Matches 72; Conservative 302; Mismatches 420; Indels 1; Gaps 1;

QY 930 TGTAACAATATCTCTGTGATCGAAAGATAAGAGTGTCTAATTGTCCTCTCTTAC 989
Db 890 HSKRRTRHHTTCRTKYNNNNNARTVYWHAAARRMMNMTTNNNNNNNNNNNNNNNACRN 831
QY 990 TCGGTCTCTATGATGACAGAAATATTCCTCAAACTCTTAATTGAGGTGATCTTC 1049
Db 830 TRTWABWKSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 771
QY 1050 TGAAGGTTCAATATTTGGAACATATCAGACACAGCTGATAAAGAGGAGTGAAGAAG 1109
Db 770 HAAVTTHTDWYKTMNTWYDMMTTMBTTTTRNNMTTSTNNNNNNNNNNNNNNNNNN 711
QY 1110 GCTGGCAATGACAACTCTATAGTTTGAAGAGGCATTTGATAAACTGAATCCTTGTC 1169
Db 710 MWKAYVAHATNNWGCWNNNTDARRTNNNTVYRRRRMMNTKTRWYSTTRRHHTYGATNN 651
QY 1170 TGCTGGAATATAGATCAGCTGATGATTCCTCAATAGTAATGAACCTCTTAAAGTAAC 1229
Db 650 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 591
QY 1230 TGAAGTGTGTACATACAGACATGACGACTTGTGTCGTGTCGTCGTCGTCGTCGTCGTC 1289
Db 590 NWMCRDVTYTRNNYCKSYAHSYVWYSSNNWYRRYSARNWSSMARWTTTRNNNNWMBV 531
QY 1290 AGTTATGTATCTGCCACAGACAGCGCCATAGTACAGCTTGTGCAAGGGGAAACACATGCT 1349
Db 530 RMRWAGTMMWRHNNNNNTDRTYVWVWKRARBTTTVYDSMNCNAKSMWRGNNWRAKMWVA 471
QY 1350 CAGAAGGTTGGCCACTCAGACAACTCCGTTGGTCATCGGAACAAAGTCACATGTTT 1409
Db 470 ANNDAGANDHWTYWGNNTHMMRRWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 411
QY 1410 GCTATATCTCATCAGCTCTCAGCTCGGTATGATCAAGATACTGATATCTCGAGGTGT 1469
Db 410 KYNWKAACNNNNBKMVRVAMWYMSRDTTNTDMMWMTSDWBWHWTVDYTMRAWNNNN 351
```

```
QY 1470 GGATTTTTCAGTCATAAATTTGGGACATATTTTCTGGAGAAATGAACATATCTTCTGTGT 1529
Db 350 NNNWBECKTTSWMMWHDHNTHTCTYGNNTWGSAYEWAAMSWAAGASNBVTYNWCRWMTY 291
QY 1530 TCATCGTGGTGAATTAATCAACTTCTAGTTCCTCACTGAAACTGTA-GTGAAGAGTAC 1588
Db 290 MGKWTNNNNNNKAWYRTKTVAWCNNRYDYDTAVMTBKRNKYCYAVBWWYBMYMGKH 231
QY 1589 AGCACTGCATCTGCTCTGTAGCCAGTACCTCAGTACGACTTCTAAGTTTGGCAGAGA 1648
Db 230 WEWRRRABHRSWNNWVCRNKYMSWHYHAMRYKNAWBAVGCNNNNWDRMAHHWCATN 171
QY 1649 AAAAATGCATAATGTTGGCATCTCTGTCACCTTTTCTTATTCCTCACTGAAATGAGGC 1708
Db 170 NNNMWWYAYMMHMKKGAATNNKTABRDDHBAHVKTYYWYRVDYWCAMCMWNAKAV 111
QY 1709 CTTCTGATGATTACC 1723
Db 110 RTAMKHMWYTTDRYV 96

RESULT 8
US-09-949-016-16919/c
; Sequence 16919, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16919
; LENGTH: 153866
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16919

Query Match 0.9%; Score 39.2; DB 4; Length 153866;
Best Local Similarity 62.0%; Pred. No. 32;
Matches 62; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 4144 ATCCGACCTGGAATAATGTCAGACATCCATGACACAGGACCAATCACTGAGTGGCT 4203
Db 20022 ATACATACTGTTGTCAGGACAGTATTAATGGAAGAGGAACTGTTCATAAAGTCGCT 19963
QY 4204 TTGTCCTCTGATGAAGATATCTTGCCACTACTCAAAACA 4243
Db 19962 GGTTCCTCAGAGGGAATAATCTGACATCTACTCAAAACA 19923

RESULT 9
US-08-956-171E-52
; Sequence 52, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
```


; INFORMATION FOR SEQ ID NO: 230:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2072 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

US-08-781-986A-230

Query Match

Best Local Similarity 0.8%; Score 38; DB 4; Length 2072;

Mismatches 80; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 3322 CCACAAATGAAAAATTTTACATCTTACGAGGAAGCGGAGCAAGCTACCGCTATT 3381

DB 1086 CCAAAATGACATAATTAATGCAATAAAGCGATAATAATTAAGTAACTTCTAACTGTT 1027

QY 3382 GTTTTACTTCGAGTAATAGAGCTGAATTTGGTGCTGAAATTCGAATTCCTAAACTATTG 3441

DB 1026 GTTGGCAACCAATATACCGCTATCGGTGGTGCTAATAATGAAGAACTGCTGTACCG 967

QY 3442 ACCAGACTCGAAGCTCTAGCCAAATTCCT 3471

DB 966 ATATTTCCCATACCAATAATACCATTTGCT 937

RESULT 14

US-10-071-411A-63

; Sequence 63, Application US/10071411A

; Patent No. 6797475

; GENERAL INFORMATION:

; APPLICANT: Joanne Barnes

; TITLE OF INVENTION: Detection of Polymorphisms in the Human

; FILE REFERENCE: MRI-021

; CURRENT APPLICATION NUMBER: US/10/071,411A

; CURRENT FILING DATE: 2002-02-07

; PRIOR APPLICATION NUMBER: 60/267,515

; PRIOR FILING DATE: 2001-02-08

; PRIOR APPLICATION NUMBER: 60/314,248

; PRIOR FILING DATE: 2001-08-21

; NUMBER OF SEQ ID NOS: 66

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 63

; LENGTH: 168174

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)...(168174)

; OTHER INFORMATION: n = A,T,C or G

US-10-071-411A-63

Query Match

Best Local Similarity 0.8%; Score 38; DB 4; Length 168174;

Mismatches 59; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 609 AATATTGAGGAGGAATCCAAACCAATTTATTGTCAGAATTGCCAAGCATCTCTTTTG 668

DB 40595 ACTATTTTGGATGATTTGTAATTAATTAATTTCTTAATACCTTTGGATTGTTATAG 40654

QY 669 TGCATTTACACAAGGTCACCTTTTGGTTGTGTGT 702

DB 40655 TTAATTTATAGAAAGGCACTGATTTTGTGTGT 40688

RESULT 15

US-10-071-411A-2

; Sequence 2, Application US/10071411A

; Patent No. 6797475

; GENERAL INFORMATION:

; APPLICANT: Joanne Barnes

; TITLE OF INVENTION: Detection of Polymorphisms in the Human

; FILE REFERENCE: MRI-021

; CURRENT APPLICATION NUMBER: US/10/071,411A

; CURRENT FILING DATE: 2002-02-07

; PRIOR APPLICATION NUMBER: 60/267,515

; PRIOR FILING DATE: 2001-02-08

; PRIOR APPLICATION NUMBER: 60/314,248

; NUMBER OF SEQ ID NOS: 66

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 63

; LENGTH: 168174

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)...(168174)

; OTHER INFORMATION: n = A,T,C or G

US-10-071-411A-63

; TITLE OF INVENTION: Detection of Polymorphisms in the Human

; FILE REFERENCE: MRI-021

; CURRENT APPLICATION NUMBER: US/10/071,411A

; CURRENT FILING DATE: 2002-02-07

; PRIOR APPLICATION NUMBER: 60/267,515

; PRIOR FILING DATE: 2001-02-08

; PRIOR APPLICATION NUMBER: 60/314,248

; PRIOR FILING DATE: 2001-08-21

; NUMBER OF SEQ ID NOS: 66

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 168273

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)...(168273)

; OTHER INFORMATION: n = A,T,C or G

US-10-071-411A-2

Query Match 0.8%; Score 38; DB 4; Length 168273;

Best Local Similarity 62.8%; Pred. No. 79;

Mismatches 59; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 609 AATATTGAGGAGGAATCCAAACCAATTTATTGTCAGAATTGCCAAGCATCTCTTTTG 668

DB 40645 ACTATTTTGGATGATTTGTAATTAATTAATTTCTTAATACCTTTGGATTGTTATAG 40704

QY 669 TGCATTTACACAAGGTCACCTTTTGGTTGTGTGT 702

DB 40705 TTAATTTATAGAAAGGCACTGATTTTGTGTGT 40738

Search completed: July 2, 2005, 13:01:55

Job time : 726 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 03:12:11 ; Search time 12549 Seconds
(without alignments)
13567.733 Million cell updates/sec

Title: US-10-645-335-1
Perfect score: 4473
Sequence: 1 atggcaggaaacagcctgt.....agcaccgcttcagtgtctaa 4473

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1843.2	41.2	2422	3	BC026618 Mus muscu
2	1279	28.6	2067	3	AK052730 Mus muscu
3	803.8	18.0	831	1	AU132411 AU132411
4	770.8	17.2	882	5	BUI188187 AGENCOURT
5	734.6	16.4	758	6	CD652969 AGENCOURT
6	670.6	15.0	678	4	B1549586 603194003
7	669.8	15.0	688	4	B1756477 603024602
8	648.4	14.5	650	5	BX507682 DKFZp686G
9	642	14.4	653	7	CN314767 170005322
10	619.6	13.9	766	4	B1855881 603383042
11	613.4	13.7	768	6	CA328908 UI-M-FY0-
12	608.2	13.6	780	5	BQ770731 UI-M-FY0-
13	601.4	13.4	603	5	BX955125 DKFZp781L
14	601.2	13.4	631	7	CR544624 DKFZp459G
15	593.2	13.3	714	7	CF535467 UI-M-GH0-
16	586	13.1	738	4	BF984064 602307513
17	577	12.9	613	1	AL042212 DKFZp434H
18	554.6	12.4	801	7	CK481829 AGENCOURT
19	553.8	12.4	693	6	BY763936 BY763936
20	546.6	12.2	878	5	BX688519 BX688519
21	544.2	12.2	625	2	BE007321 PM2-BN014
22	522	11.7	660	2	BF148762 uy34903.x
23	513	11.5	652	7	CK636646 UI-M-H00-
24	511.8	11.4	632	7	CF537838 UI-M-G10-

25	506.4	11.3	738	6	CD354970	CD354970 UI-M-GM0-
26	506.2	11.3	671	5	BM963363	BM963363 UI-M-EQ0-
27	502.4	11.2	576	2	BF043644	BF043644 BP250024A
28	497.2	11.1	517	7	CF135786	CF135786 UI-HP-BN0
29	488.8	10.9	612	7	CK636724	CK636724 UI-M-H00-
30	487.6	10.9	657	7	CNS28152	CNS28152 UI-M-HQ0-
31	479.2	10.7	659	6	CA318454	CA318454 UI-M-FY0-
32	478.4	10.7	633	6	CB525614	CB525614 UI-M-FY0-
33	476.8	10.7	653	7	CO433680	CO433680 UI-M-HX0-
34	475.6	10.6	601	6	CB582360	CB582360 AMGNNUC:N
35	475.2	10.6	628	7	CF734680	CF734680 UI-M-HB0-
36	471	10.5	601	7	CO606670	CO606670 DG8-54b23
37	459.8	10.3	817	7	CK312416	CK312416 SB02011B2
38	456.2	10.2	719	6	CD350380	CD350380 UI-M-FY0-
39	456	10.2	627	7	CO431258	CO431258 UI-M-HX0-
40	451.2	10.1	605	4	BI794834	UNN03B02
41	429.6	9.6	508	5	BQ374887	MR4-TN010
42	425	9.5	718	4	BG976355	602846632
43	420.6	9.4	531	6	CB718928	AMGNNUC:N
44	410.2	9.2	436	2	AW014238	UI-H-B10-
45	408.8	9.1	932	5	BQ713671	AGENCOURT

ALIGNMENTS

RESULT 1

BC026618

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC026618 2422 bp mRNA linear HTC 20-SEP-2002
Mus musculus, clone IMAGE:4977270, mRNA.
BC026618
GI:20071833
HTC.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Straussberg, R.
Direct Submission
Submitted (02-APR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Center Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcdpaxil@stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 59 Row: k Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Genomescan gene prediction
This clone has the following problem: frame shifted.

FEATURES

source

1..2422

/organism="Mus musculus"

/mol_type="mRNA"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:4977270"

/tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."

/clone.lib="NCI_CGAP_Mam6"
/lab_host="DH10E"
/note="Vector: pCMV-SPORTc"

ORIGIN

Query Match 41.2%; Score 1843.2; DB 3; Length 2422;

Best Local Similarity 88.4%; Pred. No. 0;

Matches 2001; Conservative 0; Mismatches 263; Indels 0; Gaps 0;

```
QY 1 ATGCGAGGAACAGCCTTGTTCTACCCATTGTTCTTTGGGTCGAAAGCGCCACACAT 60
DB |||
DB 158 ATGCGAGGAACAGCCTTAGTCTCGCCATTGTTCTTTGGGCGCGCAAGCACCCACACAT 217
QY 61 TGCATCTCAGCGTACTTTTAAACAGATGATGGCGCCACAGATCGTAAACAGATGTCAGGAC 120
DB |||
DB 218 TGCATTTGTCGAATCTGTTGACAGATGATGGGGGACAAATGTAATGCGATGCCACGAT 277
QY 121 GGACAAATATGTCCTCGGGATCTTTTCAGTAGAACTGCAAAATTAATCCTCGAGCACTGTTG 180
DB |||
DB 278 GGACAAATATGTCCTCGGGATGTTTCGGTAGAACTAGNAGTTAATCCCCGAGCACTGTTA 337
QY 181 TTTGGTCATACAGCATCAATCACTTGTTTGTCTAAAGCTTGTGCTTCCAGTGACAAACAG 240
DB |||
DB 338 TTTGGCCACACAGCATCCATCACTTGTTGTCAAAAGCCTCGCCTTCTGGGACAAAGCGG 397
QY 241 TATATTGTAGTGCATCTGAAAGTGGAGATGTCCTCTGGATGCTGAGTGATGGCAGA 300
DB |||
DB 398 TACACTGTGAGCGCGTCTGCAACCGAGAGATGTGCCCTCTGGATGTGAACGATGGCAGA 457
QY 301 TGTATTGAAATTTACAAAATTTAGCTTGACACATCTGGCATACAGTTCTACCAAGTTCTCT 360
DB |||
DB 458 TGTATTGAAATTTACAAATTTAGCTTGACACATCTGGCATACAGTTCTACCAAGTTCTCT 517
QY 361 GTTGGGAATCAGCGAAGAAAGGCTTTTATGCGACGACATTAACCTGGAATTCCTTGT 420
DB |||
DB 518 GTTGGGAACAGCAAGAGGCGAGGCTCCTCTGCCATGGACATTAACCTGGAATTCCTCGTT 577
QY 421 GTGGATGCTACAGCGCTTGAAGTATTATCTCTCTAGTATCAAGATATACCAAGACTGG 480
DB |||
DB 578 GTGGATGCCACAGCGCTTGAAGTGTGATATCTCTGTTGTAATCTGGAATCGAAGATCTCTCAGACTGG 637
QY 481 ATTAGCTCCATGAGTATTAATTCGATCCACCGCAACACAGAGGACACAGTGGTAGCACTC 540
DB |||
DB 638 ATCAGCTCCATGAGCATCATCCGCTCTCAGCGGACACAGAGGACACTGTGTGGCGCTC 697
QY 541 TCGGTGATCGGCATCTCGAAGGCTGCGATTGTTTACCTCGGAATAAATGACATGCGAGGAT 600
DB |||
DB 698 TCTGTGACAGGTATTCTGAAGGTGTGGATTGTCAACCTCTGAAATGAGTGGAAATGCAGGAT 757
QY 601 ACTGAGCCATATTTGAGGGGAATCCAAACCAATTTATGTCAGAAATGCGCAAGCATC 660
DB |||
DB 758 ACTGAGCCAAATTTGAGGGGAATCCAAACCAATTTATGTCAGAAATGCGCAAGCATC 817
QY 661 TCTTTTGTGCAATTTACAAAGGTCACTTTTGGTTGTGTTTCCAAATATTGAGGGGTG 720
DB |||
DB 818 TCTTTTGTGCAATTTACAAAGGTCACTTTTGGTGTGATGCTCCAAATACTGAGGGGTG 877
QY 721 TTCGATCGCGGACATTAATCTCTTGTGTGTTTCAAGTCCCTAGTGAAATGAGCAGACATGG 780
DB |||
DB 878 TTTGATGCTGGGCACTACTCTGTGTGCTCAGGTCCTAGTCCAGTCCAGTGGACAGACATGG 937
QY 781 ACCGGGGGACATTTGCTCATCAGATAAGTCCATCATTTGGACAGAAATGGGCAAGT 840
DB |||
DB 938 ACTGGAGGGGACATTTGTGTCTGACAGAAAGTCAATCATCTGGACAGAAATGGGCAAGT 997
QY 841 TATATTTACAAACTACTCGCCAGTTGCCCTTCCAGCTAGTGAATCAATTCGCCAGTGTG 900
DB |||
DB 998 TACATCTACAACTCCCTGCGAGTTGCCCTTCCAGCTAGTGAATCAATTCGCCAGGACGTG 1057
QY 901 GGGAGGCGAGTTGAAATTTAATTCCTCTGTTTCAACATATTCCTTTGGATCGAAAGAT 960
DB |||
DB 1058 GGGAAAGCAGTGGAAAAATTTGATCCCTCTGTCAGCATAGCCTCTTTGGATCAGAAAGAT 1117
```

```
QY 961 AAAGAGTTGCTAATTTGTTCTCTCTGTTACTCGGTTCTTCTATGATGACAGAAATATTTTC 1020
DB |||
DB 1118 AAAGAGTTGTTAATTTGTTCTCTCTGTTACTCGGTTTCTACGATGCAAGGAATATTG 1177
QY 1021 CATAAACTGTTAATTTACGGGTGATTTCTTGGAGGTTGAAATATTTTGGAAACATATCAGAC 1080
DB |||
DB 1178 CATAGCTACTAATTTAGGGCGATTCTTCTGGAAGGTTAAATATCTTGGAAACATAGCAGAC 1237
QY 1081 ACAGCTGATAAACAGGGAAGTGAAGAGGCTGGCAATGACAACTTCTATTAGTTTGGCAA 1140
DB |||
DB 1238 ATAGCAGAGAAACAGGAAGCCGATGAAGGGCTAAAGATGACAACTTGTATTAGCTTGGCAA 1297
QY 1141 GAGGCATTTGATAAACTGAATCCTTGTCTCTGCTGGAATTTATAGATCAGCTGAGTGTGATT 1200
DB |||
DB 1298 GAGGCATTTGACAGCTGAAGCCTTGGCCTCTGGBAATTTATCGATCAGCTAAGTGTGATT 1357
QY 1201 CCCAATAGTAAATGAACCTCTTAAAGTAACCTGCAAGTGTGTATACACAGACATGACGA 1260
DB |||
DB 1358 CCCAACAGCAACGAACTCTTAAAGTAACCTGCGAGTGTCTACATACAGCACACGCGCGA 1417
QY 1261 CTGTTTGTGCTCGTGAAGATGGAAGCATAGTTATTGTACCTGCCACACAGACGGCCATA 1320
DB |||
DB 1418 CTGTTTGTGCGCGGGAAGATGGAAGCATCATTTATTGTGCTGCCACCCAGACGGCCATA 1477
QY 1321 GTACAGCTGTGCAAGGGGAACAACATGCTCAGAGAGGTTGGCCACTCAGACAAACACTC 1380
DB |||
DB 1478 GTTACGTATTGCAAGGAGAACACATGCTCAGAGAGGTTGGCCCTCAGACAAACCTC 1537
QY 1381 CGTGTTCATCGGAACAAAGTCACTGTTTGTCTATATCTCATCAGGTCTCAGCTCGGTAT 1440
DB |||
DB 1538 CGTGGCCATCGGAAACAAAGTCACTGCTGCTGTATCCCCATCAGGTCTCAGCTCGGTAT 1597
QY 1441 GATCAAGATACCTGATATCTGAGGTGTGGATTTTTCAGTCAATATTTTGGGACATATTT 1500
DB |||
DB 1598 GACCAAGATACCTGATATCTGAGGTGTGGATTTTTCGTCATATTTTGGGACATATTT 1657
QY 1501 TCTGGAAGAAATGAAACATATCTTCTGTGTTTCATGTTGTGGTGCAGATTACTCAACTCTAGTT 1560
DB |||
DB 1658 TCTGGGAATGGAACATATCTTCTGTGTTTCATGTTGTGGTGGTGGATCACACAACTCTTGGT 1717
QY 1561 CCACCTGAAATCTGTAGTGCAGAGTACAGCATGTCATCTGCTCTGTAGCGATGACCCAC 1620
DB |||
DB 1718 CCCCCGAAATCTGTAGTGCAGAGTTCACACCTGTCATCTGTTCCGTAGCCAGTGCAC 1777
QY 1621 TCAGTAGGACTTCTAAGTTTGGAGAGAAATGCATATGTTGGCATCTGTCACCTT 1680
DB |||
DB 1778 TCTGTAGGCTCTCTAAGTCTGCGAGAGAAAAATGCATCATGTTGGCGTCTGCCACCT 1837
QY 1681 TTTCTATTCAAGTAATCAATGGAGGCTTCTGATGATTAACCTGTTGGTGGGCTGTTC 1740
DB |||
DB 1838 TTTCTATTCAAGTGTATCAAGTGGAGGCTTCCGACGACTACCTGTTGGTGGGATGCACA 1897
QY 1741 GATGTTCTGTGTACGTTGGCAATGGAATCTGTTGCAATTTGGATCGTTGTGTGATGGG 1800
DB |||
DB 1898 GACGGCTCTGTGTATGCTGTGGCAGATGGAACACTGTTGCGCTGGATCGTTGTGCAATGGGA 1957
QY 1801 ATAAACAGCATTTGAGATTTCTAAACGCTTGTGATGAGCTGTTCCCTGCTGCTGTGATTC 1860
DB |||
DB 1958 ATCACAGCATGAGGATACCTGAAACGCTTGTGACGAAGCGGTTCTCCCGCTGTGAGACTCT 2017
QY 1861 CTTAGTTCATCCAGCAGTCAACCTTAAACCAAGCTATGACGAGACGTAGTCTTGTGCTCTT 1920
DB |||
DB 2018 CTTAGTTCATCCAGCAGTCAACCTGAAACCAAGCCATGACAAAGGAGTCTCGCGCGCTT 2077
QY 1921 AAAAATATGGCCCATCATTAAGCTACAAACCTTTGCAACTTAACTCTTTGGCTTCTGAGGCA 1980
DB |||
DB 2078 AAAAATATGGCCCAACATTAAGCTGCAAAACCTTTGCAACGAAACCTTTTGGCTTCTGAGGCC 2137
QY 1981 TCTGCAAGGGGAAATTTTACATAATTTCTCAATCTCCCTGATGTTTCAAGCAATAAAG 2040
DB |||
DB 2138 TCTGCAAGGGGAAATTTTACATAATTTCCCAATCTCCCTGATGTTTCAAGCAATAAAG 2197
QY 2041 ACAAACCTTAACAGACCCGACACATACATGTGCTATTCTTTGATGTGGAGCGTTGATTAT 2100
```

```

Db      2198  ACAAACCTACAGACCGGATATCCAGTCTTTCTTTGATGTGAAGCTTGATATT 2257
QY      2101  CAACTCCTGACTGAAGAAGCCTTAGGCCGAATACCTGCTCTTATTTCCCGAGAGATTTG 2160
Db      2258  CAACTCCTGACTGAAGAAGCCTTAGGCCGAATACCTGCTCTTATTTCCCGAGAGATCTG 2317
QY      2161  CAAAAGACATCGGACGTTTCAGACAAAGGGGCTCTTTTAACTGTGAAACGAGCAGCA 2220
Db      2318  CAGAAGACATCTGGCAGTTTCAGACAAAGGGGCTCTTTCTTAACGTGGGAAGCAGCAGCG 2377
QY      2221  GTTCTCTCCCAACAGTGAAGAAACGATCAAGACAAACATCAA 2264
Db      2378  GTTCTTTTCAGCAAGTGAAGAAACATCAAAAAAATAA 2421

```

RESULT 2

AK052730

LOCUS

DEFINITION

AK052730 Mus musculus 0 day neonate kidney cDNA, RIKEN full-length enriched library, clone:D630037A10 product:hypothetical G-protein beta WD-40 repeats containing protein, full insert sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

DEFINITION 603194003F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5265024 5',
mRNA sequence.
ACCESSION B1549586
VERSION B1549586.1 GI:15436898
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NIH-MGC http://mgc.nci.nih.gov/
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov.
Plate: LLAM11668 row: f column: 01
High quality sequence stop: 678.
Location/Qualifiers
1. .678
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5265024"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC 95"
/notes="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site: 1: BamHI; Site 2: SalI-XhoI
(gtagag); Oligo-dT primed using primer
5'-TTTTTTTTTTTTTTVN-3', size-selected for average
insert size 2.5 kb and normalized to ROT 5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NIH/NHGRI, National Institutes of Health). Note: this
is a NIH_MGC Library."

ORIGIN
Query Match 15.0%; Score 670.6; DB 4; Length 678;
Best Local Similarity 99.4%; Pred. No. 3.3e-182;
Matches 673; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2625 AGGAAAGGAACTTACGGAGTGTCCCGTCCGTCACACAGCATCTCCTGTCTATCAT 2684
DB 1 AGCGAGGGAACCTTACGGAGTGTCCCGTCCGTCACACAGCATCTCCTGTCTATCAT 60
QY 2685 TTCTTTGGCAATACCTTTAATGAGTAGTGACCAATGCAACTTTTATTGGTGATCATATGAA 2744
DB 61 TTCTTTGGCAATACCTTTAATGAGTAGTGACCAATGCAACTTTTATTGGTGATCATATGAA 120
QY 2745 GAAGGTCTCTACAGGCCACTAGACCAACACCCAGACCTTCTAAGCAAGGGGTTTC 2804
DB 121 GAAGGTCTCTACAGGCCACTAGACCAACACCCAGACCTTCTAAGCAAGGGGTTTC 180
QY 2805 CCTCCAACTTCAGTAATATTGCAAGGACAGATTAAACAAGTTGCTGCACCTGTCTGT 2864
DB 181 CCTCCAACTTCAGTAATATTGCAAGGACAGATTAAACAAGTTGCTGCACCTGTCTGT 240
QY 2865 TTCGGTCTGTGATCTGATCACTCTGGCTCTGACCCCTCTCTCTCTCTCTCTTACA 2924
DB 241 TTCGGTCTGTGATCTGATCACTCTGGCTCTGACCCCTCTCTCTCTCTCTTACA 300
QY 2925 TACTGTCTTTCTAGTAATGAAGTTGGAGTGTAGCTAGCTATGACATCTGTGTATGCT 2984
DB 301 TACTGTCTTTCTAGTAATGAAGTTGGAGTGTAGCTAGCTATGACATCTGTGTATGCT 360

QY 2985 GCAGAGCTACTCGGATTTGGATTAATTTAGGCTCCCTCTCTGAGATGCTGCCCGAAG 3044
DB 361 GCAGAGCTACTCGGATTTGGATTAATTTAGGCTCCCTCTCTGAGATGCTGCCCGAAG 420
QY 3045 ATGGCAAGATCGATGCTTTGGAGGTGAGAGAAGCCGACAGGCCCTGCTTCTGGCGGAAC 3104
DB 421 ATGGCAAGATCGATGCTTTGGAGGTGAGAGAAGCCGACAGGCCCTGCTTCTGGCGGAAC 480
QY 3105 GAGAAGAATTTGAGCAGGCGAGGAGGAAGCAATTTGATGCTGGGCTCTTTACTTTACC 3164
DB 481 GAGAAGAATTTGAGCAGGCGAGGAGGAAGCAATTTGATGCTGGGCTCTTTACTTTACC 540
QY 3165 TCAGTACATAGACCACTGTCATATCAGCTGAGGTGATCATCAGAGAAGCCGCGAGACTATCAC 3224
DB 541 TCAGTACATAGACCACTGTCATATCAGCTGAGGTGATCATCAGAGAAGCCGCGAGACTATCAC 600
QY 3225 CACGGCTCTGTGCTCAGGCGCTCAGGCGCTGAGCAAAAGTCCAGGAGAGAGCATGACCTTGT 3284
DB 601 CACGGCTCTGTGCTCAGGCGCTGAGCAAAAGTCCAGGAGAGAGCATGACCTTGT 660
QY 3285 TGACGATGACATCACCA 3301
DB 661 TGACGATGACATCACCA 677
RESULT 7
BI756477 688 bp mRNA linear EST 25-SEP-2001
LOCUS 603024602F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5195065 5',
mRNA sequence.
ACCESSION BI756477
VERSION BI756477.1 GI:15748055
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NIH-MGC http://mgc.nci.nih.gov/
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11488 row: c column: 02
High quality sequence start: 3
High quality sequence stop: 688.
Location/Qualifiers
1. .688
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5195065"
/lab_host="DH10B"
/clone_lib="NIH_MGC 114"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: EcoRV (destroyed); RNA source anonymous pool of 6
male brains, age range 23-27 yo. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 019. Note:
this is a NIH_MGC Library."

FEATURES
source

ORIGIN

```
Query Match      15.0%; Score 669.8; DB 4; Length 688;
Best Local Similarity 99.6%; Pred. No. 5.6e-182;
Matches 682; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 3668 TGGGGCTTCTCGAATCTTTGTGCGGATGCGGAGAAACAATTCGCAACATCACAATGGGGT 3727
Db 1 TGGGGCTTCTCGAATCTTTGTGCGGATGCGGAGAAACAATTCGCAACATCACAATGGGGT 60

QY 3728 TGCCTCTGAGCCAGCAGCTGCTCGGCGCGCTCTCGGAGGATGCGCTCTGCTCATG 3787
Db 61 TGCCTCTGAGCCAGCAGCTGCTCGGCGCGCTCTCGGAGGATGCGCTCTGCTCATG 120

QY 3788 CCACCGCCAGACCAACCGCGCTTCATCACCACCATAGCCAAAGAGGTACACAGATACCG 3847
Db 121 CCACCGCCAGACCAACCGCGCTTCATCACCACCATAGCCAAAGAGGTACACAGATACCG 180

QY 3848 CTCTTCGAGCAATACCCCAATCAGCAGAAATGACACAACTCTTTGACGAGCTA 3907
Db 181 CTCTTCGAGCAATACCCCAATCAGCAGAAATGACACAACTCTTTGACGAGCTA 240

QY 3908 AAGGGAAATTTGAGAGTCATTGAATTTCTATTGAAAGATGCCACAGATGTTGG 3967
Db 241 AAGGGAAATTTGAGAGTCATTGAATTTCTATTGAAAGATGCCACAGATGTTGG 300

QY 3968 ATCTTCTGAGGATTTATGGACATCATATTATGCTGCTTGAAGATCTTTAGTTAAA 4027
Db 301 ATCTTCTGAGGATTTATGGACATCATATTATGCTGCTTGAAGATCTTTAGTTAAA 360

QY 4028 AGAAGGCTCTCAAGATGTTTCCAGCCATCTGCAAGGTTCTACATGTCAGCTATTATG 4087
Db 361 AGAAGGCTCTCAAGATGTTTCCAGCCATCTGCAAGGTTCTACATGTCAGCTATTATG 420

QY 4088 AGCGGAATCACAGATAGCAGTTGGAGCTGCGCATGTTTCACTGAGGCTGACGATCC 4147
Db 421 AGCGGAATCACAGATAGCAGTTGGAGCTGCGCATGTTTCACTGAGGCTGACGATCC 480

QY 4148 GGACTGGAATGTCAGACAAATCCATGGACACAAAGGACCAATCACTGAGTGGCTTTG 4207
Db 481 GGACTGGAATGTCAGACAAATCCATGGACACAAAGGACCAATCACTGAGTGGCTTTG 540

QY 4208 CTCCTGATGGAAGATATTGCAACCTACTTCAAACTGACAGCCACATTTCTTTTGGC 4267
Db 541 CTCCTGATGGAAGATATTGCAACCTACTTCAAACTGACAGCCACATTTCTTTTGGC 600

QY 4268 AGATGAACAGTCACCTGCTGGGAGCATGG-CATGCTGAATCGGCACTCAGCTGCCG 4326
Db 601 AGATGAACAGTCACCTGCTGGGAGCATGGCCATGCTGAACCTCGGCACCTCAGCTGCCG 660

QY 4327 TGCAATTAACACCTACCAGGTGCCCC 4351
Db 661 TGCAATTAACACCTACCAGGTGCCCC 685
```

```
RESULT 8
BX507682      650 bp mRNA linear EST 04-SEP-2003
LOCUS      DKFp686G03261_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DEFINITION      DKFp686G03261 5', mRNA sequence.
ACCESSION      BX507682
VERSION      BX507682.1 GI:32040573
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 650)
AUTHORS      Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M. and
Wiemann S.
TITLE      EST (Poustka A., Albert R., Moosmayer P., Schupp I.,
Wellenreuther R., et al.)
JOURNAL      Unpublished (2003)
COMMENT      Contact: MIPS
```

```
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFp686G03261) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
Location/Qualifiers
1..650
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFp686G03261"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplex2; Site_1: SfiI; Site_2: SfiIB;
cDNA-collection"
```

ORIGIN

```
Query Match      14.5%; Score 648.4; DB 5; Length 650;
Best Local Similarity 99.8%; Pred. No. 8.8e-176;
Matches 649; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3062 TGGAGGTGAGAGAACCGCACAGGCCCTGCTTCTGGCGAAGTGAAGAATTTGACAGG 3121
Db 1 TGGAGGTGAGAGAACCGCACAGGCCCTGCTTCTGGCGAAGTGAAGAATTTGACAGG 60

QY 3122 CAGGACAGAGAGAACCATTTGCTGGGCTTCTTACTTACCTCAGTACATAGACCAG 3181
Db 61 CAGGACAGAGAGAACCATTTGCTGGGCTTCTTACTTACCTCAGTACATAGACCAG 120

QY 3182 TCATATACCTGGAGTCATCAGAAAGCCGCGAGACTATCACCAGGCTCTCTGATGCT 3241
Db 121 TCATATACCTGGAGTCATCAGAAAGCCGCGAGACTATCACCAGGCTCTCTGATGCT 180

QY 3242 CAGGCGCTGAACAAAGTCCAGAGAGAGAGCATGACCTTTTGACGATGATCACCAC 3301
Db 181 CAGGCGCTGAACAAAGTCCAGAGAGAGAGCATGACCTTTTGACGATGATCACCAC 240

QY 3302 CTGGTTGCTTATCAAGTGTCCACAAAATGAAAAAATTTCTACATCTTACGAGGAAGAC 3361
Db 241 CTGGTTGCTTATCAAGTGTCCACAAAATGAAAAAATTTCTACATCTTACGAGGAAGAC 300

QY 3362 GGAGCAGCTACCGCTATTGTTTACTTGGAGTAAATAGGAGCTGAATTTGGTGTGAAA 3421
Db 301 GGAGCAGCTACCGCTATTGTTTACTTGGAGTAAATAGGAGCTGAATTTGGTGTGAAA 360

QY 3422 TTGAACCTCTTAAACTATTGACAGACCTCGAAGCTTAGCCAAATTTCTGAGGGATTG 3481
Db 361 TTGAACCTCTTAAACTATTGACAGACCTCGAAGCTTAGCCAAATTTCTGAGGGATTG 420

QY 3482 GGTGACTAGTGGTGGATCCAACTACTGCTGGCCAGACATACTTGCAGGCACTGACGT 3541
Db 421 GGTGACTAGTGGTGGATCCAACTACTGCTGGCCAGACATACTTGCAGGCACTGACGT 480

QY 3542 TTCTTCTGCTACAGCTCCAGGCCCAACTTCTCCACAGAGCACTATCCGAAGAACAG 3601
Db 481 TTCTTCTGCTACAGCTCCAGGCCCAACTTCTCCACAGAGCACTATCCGAAGAACAG 540

QY 3602 CCATTGATCTGATGGAAGTGGGTTTCACTGTTTGGAGGCTTTACATGATGTTGTCGCTG 3661
Db 541 CCATTGATCTGATGGAAGTGGGTTTCACTGTTTGGAGGCTTTACATGATGTTGTCGCTG 600

QY 3662 TTCTGATGGGCTTCTCGAACTTTGTGCGGATGCGCGAGAAAACAATTGGC 3711
Db 601 TTCTGATGGGCTTCTCGAACTTTGTGCGGATGCGCGAGAAAACAATTGGC 650
```

```
RESULT 9
CN314767
LOCUS
DEFINITION 1700532202315 GRN_ES Homo sapiens cDNA 5', mRNA linear EST 16-MAY-2004
ACCESSION CN314767
VERSION CN314767.1 GI:47331181
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 653)
Li, Y., Xu, C., Fang, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G. J.,
Lebkowski, J and Stanton, L. W.
Transcriptome characterization elucidates signaling networks that
control human ES cell growth and differentiation
Nat. Biotechnol. 22 (6), 707-716 (2004)
JOURNAL Contact: Brandenberger R
COMMENT Regenerative Medicine
Geront Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@geron.com
Insert length: 653 Std Error: 0.00.
FEATURES
Location/Qualifiers
source
1..653
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cells, cell lines H1, H7, and H9"
/clone_lib="GRN ES"
/notes="oligo dt primed, full-length enriched cDNA library from undifferentiated hES cell lines H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free conditions"
ORIGIN
Query Match 14.4%; Score 642; DB 7; Length 653;
Best Local Similarity 99.8%; Pred. No. 6.4e-174;
Matches 653; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 3449 CTCGAAGCTCTAGCCAAATTCCTGAGGGATTCGGGTGACTAGTGGTGGATCCAACTACT 3508
Db 1 CTCGAAGCTCTAGCC-AAATTCCTGAGGGATTCGGGTGACTAGTGGTGGATCCAACTACT 59
QY 3509 CGCTGGCCGACACATCTTGCAGGCACTGACGTTTCTTCTGCTACAGCTCCAGCCCCA 3568
Db 60 CGCTGGCCGACACATCTTGCAGGCACTGACGTTTCTTCTGCTACAGCTCCAGCCCCA 119
QY 3569 AACTCTCTCCACACAGCACTATCCGAGACACCCATTGATCTGATGGCGGTCTCA 3628
Db 120 AACTCTCTCCACACAGCACTATCCGAGACACCCATTGATCTGATGGCGGTCTCA 179
QY 3629 CTGTTTGGAGCGCTTACATGGATGTGTCGCTGTCTTGATGGCGGTCTTCTGAACTTTGG 3688
Db 180 CTGTTTGGAGCGCTTACATGGATGTGTCGCTGTCTTGATGGCGGTCTTCTGAACTTTGG 239
QY 3689 CCGATCCGAGAAACAACTTGGCCAAACATCAATCAATGGGTTGCTCTGAGCCGACGCTG 3748
Db 240 CCGATCCGAGAAACAACTTGGCCAAACATCAATCAATGGGTTGCTCTGAGCCGACGCTG 299
QY 3749 ACTCGGCCGCTCTGGGGCATGCCCTCTCGCTCATTCGCCCGCCAGACACCGCCT 3808
Db 300 ACTCGGCCGCTCTGGGGCATGCCCTCTCGCTCATTCGCCCGCCAGACACCGCCT 359
QY 3809 TCATCACCCATAGCAAGAGGTACACAGATACGGCTCTTTCGAGCAAAATACCCAAAT 3868
Db 360 TCATCACCCATAGCAAGAGGTACACAGATACGGCTCTTTCGAGCAAAATACCCAAAT 419
QY 3869 CACAGCAGAAATATGCACACAACTCTTTCACAGACTAAAGGGGAAATTTTGAGAGTCA 3928
Db 420 CACAGCAGAAATATGCACACAACTCTTTCACAGACTAAAGGGGAAATTTTGAGAGTCA 479
QY 3929 TTGAAATTTCTTATTTGAAAGATGCCACAGATGTTCTGAGTCTTCTCGTGGAGTTATGG 3988
Db 480 TTGAAATTTCTTATTTGAAAGATGCCACAGATGTTCTGAGTCTTCTCGTGGAGTTATGG 539
QY 3989 ACATCATTTATGTTACTGCTTGAAGGATCTTTAGTAAAGAAAGGTCTTCAAGATGTT 4048
Db 540 ACATCATTTATGTTACTGCTTGAAGGATCTTTAGTAAAGAAAGGTCTTCAAGATGTT 599
QY 4049 TCCACGCCATCTGCAGGTTCTACATGCTCAGCTATTATGAGCGGAATCACAGAA 4102
Db 600 TCCACGCCATCTGCAGGTTCTACATGCTCAGCTATTATGAGCGGAATCACAGAA 653
RESULT 10
BI855881
LOCUS
DEFINITION 603383042F2 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5391706 5', mRNA linear EST 10-OCT-2001
ACCESSION BI855881
VERSION BI855881.1 GI:15996628
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 766)
NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Prepared by: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLML1998 row: d column: 11
High quality sequence stop: 719.
FEATURES
Location/Qualifiers
source
1..766
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5391706"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="PH10B (phage-resistant)"
/clone_lib="NIH MGC 87"
/notes="Organ: breast; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.383 Kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
ORIGIN
Query Match 13.9%; Score 619.6; DB 4; Length 766;
Best Local Similarity 94.9%; Pred. No. 2.2e-167;
Matches 729; Conservative 0; Mismatches 24; Indels 15; Gaps 8;
QY 3294 CATCACCCTGG-TTGCTTATCAAG-TGTCACCAAAATCAAAAAATTTCTACATCTTAC 3351
Db 4 CATCACCCTGGATTGCTTATCAGATGTCCCAAAATGAGAAATTTCTACATCTTAC 63
QY 3352 GAGGAA-AGACGGAAGCAAGCTACCGCTATTGTTTCTTGGAGTAATAGGAGCTGAAT 3410
Db 64 GAGGAAATAGCAGGAAGCAAGCTACCGCTATTGTTTCTTGGAGTAATAGGAGCTGAAT 123
QY 3411 TGGTGTGAAATTTGAACCTCTTAACTATTATGACGACCTCGAAGCTCTAGCCAAATTC 3470
Db 3470 TGGTGTGAAATTTGAACCTCTTAACTATTATGACGACCTCGAAGCTCTAGCCAAATTC 3470
```

```
Db 124 GGTGTC-----TGAACCTCTAAACTATATAGACGACGACCTCGAAGCTCTAGCCAAATACC 177
QY 3471 TAGGGGATTCGGGTGACCTAGTGGTGGATCCAACTACTCTGCTGGCCAGACATACCTTGCAA 3530
Db 178 TAGGGGATTCGGGTGACCTAGTGGTGGATCCAACTACTCTGCTGGCCAGACATACCTTGCAA 237
QY 3531 GGCACCTAGCTTCTCTCTACTAGCTCAGGCTCCAGGCCCAAACTTCCTCCACAGCACTAT 3590
Db 238 GGCACCTAGCTTCTCTCTACTAGCTCAGGCTCCAGGCCCAAACTTCCTCCACAGCACTAT 297
QY 3591 CCGAAGAACAGCCATTCATTCGATTCGAGCTGGGTTCACCTGTTTGGGAGCCTTACATGGA 3650
Db 298 CCGAAGAACAGCCATTCATTCGATTCGAGCTGGGTTCACCTGTTTGGGAGCCTTACATGGA 357
QY 3651 TGTGTCGGCTGTTCTCTGATGGGGTTCCTGAACTTTTGTGCGGATGCGCGAGAAACAATTGC 3710
Db 358 TGTGTCGGCTGTTCTCTGATGGGGTTCCTGAACTTTTGTGCGGATGCGCGAGAAACAATTGC 417
QY 3711 CAACATCACAAATGGGCTTGCCTTGAGCCGACGAGCTGACTCGGCCGCTCTGCGAGGCA 3770
Db 418 CAACATCACAAATGGGCTTGCCTTGAGCCGACGAGCTGACTCGGCCGCTCTGCGAGGCA 477
QY 3771 TGCCTCTCTGCTCATTTGCCACCCGACAGACCAACCGGCTTCATCACCAATAGCCAAAGA 3830
Db 478 TGCCTCTCTGCTCATTTGCCACCCGACAGACCAACCGGCTTCATCACCAATAGCCAAAGA 537
QY 3831 GGTACACAGACATACGGCTCTTTCAGCAATATCCCAATCACAGCAGAAATGACACAAC 3890
Db 538 GGTACACAGACATACGGCTCTTTCAGCAATATCCCAATCACAGCAGAAATGACACAAC 597
QY 3891 AACTCTTGACGAGCTTAA--GGGAAATTTTCAGAGTCAATGAAATTTCTTATGGA-AAA 3947
Db 598 AACTCTTGACGAGCTTAAAGGGGACAAATCTGAGAGTCATGAAATTTCTTATGACAAA 657
QY 3948 GATGCCCAACAGATGTTGT--GGATTTCTCTGTGAGGTTATGGAATCATTTATGATGTC 4006
Db 658 GATGCCCAACAGATGTTGTGGATTTCTCTGTGGAAGATGAGACATCAATATGATGTC 717
QY 4007 TTGAA--GGATCTTTAGTTAAAGAAAGGCTTTCAGAAATGTTTCCC 4052
Db 718 TTGAAAGGATCTTAGTTAAACAGAAAGCTTCTCAAGAAATGTTTCCC 765

RESULT 11
CA328908
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 768)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: gcapbs-1@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1. .768
source
```

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6827856"
/tissue_type="whole brain"
/dev_stages="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_FY0"
/notes="Organ: Brain; Vector: pYX-Asc; Site:1: EcoR I;
Site:2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is ACGGAGACAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

ORIGIN

Query Match      13.7%; Score 613.4; DB 6; Length 768;
Best Local Similarity 87.4%; Pred. No. 1.4e-165;
Matches 671; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 1775 GTGCATTGGATCGTTGTGTGATGGGATAACAGCAGTTCAGAGATTCCTAAACGGTTGTGTG 1834
Db 1 GTGCGCTGGATCGTTGTGCAATGGGAAATCACAGCAGTGGAGATCTGAACGCTTGTGACG 60

QY 1835 AAGCTGTTCTCTGCTGCTGTGTTGATTCATCTTAGTTCATCCAGAGTCAACCTTAAACAAGCTA 1894
Db 61 AAGCGGTTCTCTCGCTGTAGACTCTCTTAGTTCATCCAGCAGTCAACCTTGAACAAGCCA 120

QY 1895 TGACGAGAGTGTCTTGTGCTCTTTAAATAATGCCCCATCATAGCTTACAAACCGTTG 1954
Db 121 TGACAAGACGAGTCTCGCGCGCTTTAAAAAATATGCCCCCACCATAGCTGCAAAACCGTTG 180

QY 1955 CAACTAAACCTCTTGGCTCTGAGGATCTGACAAGGGAAATTTACCTAAATATTTCTCATA 2014
Db 181 CAAAGAACCTTTTGGCTCTGAGGCTCTGACAAGGGTAATTTACTTAATATTTCCATA 240

QY 2015 ACTCCCTGATGTTCAAGCAATAAAGACAAACCTTAACAGACCCGAGACATACATGTGCTAT 2074
Db 241 ACTCCCTGATGTTTCAAGCAATAAAGACAAACCTTAACAGACCCGAGATATCCATGTGCTTT 300

QY 2075 TCTTTGATGTGAGAGGTTGATTTTCACTCTGACTGAAAGGCTCTAGGCCGAATA 2134
Db 301 TCTTTGATGTGAGAGGTTGATTTTCACTCTGACTGAAAGGCTCTAGGCCGAATA 360

QY 2135 CTGCTCTTATTTCCCGAGAGATTTTGCACAAAGCATCTGCCAGTTTCAGACAAAGGGGCT 2194
Db 361 CTGCACTTATTTCCCGAGAGATTCGACAAAGCATCTGGCAGTTTCAGACAAAGGGGCT 420

QY 2195 CTTTFTTAACTGGAAAACGAGCAGCTGTTCTTTCCAAACAAGTGAAGAAACGATCAAG 2254
Db 421 CTTTCTTAACTGGGAAGCGAGCAGCGGTTCTTTTCCAGCAAGTGAAGAACTATCAAG 480

QY 2255 AGNACATCAAGNACACCTCTTGTGATGAGAGGAGGATGAGGAGATTAATGAGGCAGA 2314
Db 481 AGAACAATAAAGGAACACCTTCTGGATTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540

QY 2315 GAAGGGAAGAAAGTGTATCTGATATCGGTTCAGCAAAATCAAAAGCATTTGACCTATTAG 2374
Db 541 GCNAGAGGATAGTAGCCCTCTGACTACCGGCCAGCAAGTCCAGGCCACTGACCTACTAG 600

QY 2375 AATATAATTTAACTATGGACACTGTGCAAGAGCTGTTTATGCTCTCGCTTCAGGCTGGGTT 2434
Db 601 AATACAACCTTACTATGGATACCGCAANAATTTATTCATGCTCTCTCCAGCGCTGGGCT 660
```


QY 2435 TGAATGAAGTACTGGATGAAGTTTGGCTTGGATCGCCTTGGAACTGCAAAACCCCACTGCA 2494
 DB 661 TGAATGAAGTACTGGATGAAGTTTGGCTTGGATCGCCTGGAATGCTGAAACCGCACTGCA 720
 QY 2495 CCGTATCGTTGGCTCTTGTGACAGAGGAGGCCATATGTCACATGATCC 2542
 DB 721 CGGTGCTCTTGGCTCTCTATCGAGAGAGGTCTATATGTCGCTGATGC 768

RESULT 12
 B0770731
 LOCUS
 DEFINITION UI-M-F10-byw-a-03-0-UI.r1 NIH BMAP_F10 Mus musculus cDNA clone
 IMAGE:6400010 5', mRNA sequence.
 B0770731
 B0770731.1 GI:21979205
 EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM
 Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 780)
 NIH-MGC <http://mgi.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pYX-5.
 Location/Qualifiers
 1..780
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:6400010"
 /tissue_type="whole brain"
 /dev_stage="embryo 12.5dpc"
 /lab_host="DH10B (r1 phage resistant)"
 /clone_lib="NIH_BMAP_F10"
 /notes="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is CAGCCAGAC. This library was created for the University
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 program coordinator."

ORIGIN

Query Match 13.6%; Score 608.2; DB 5; Length 780;
 Best Local Similarity 88.1%; Pred. No. 4.4e-164;
 Matches 661; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 3724 GGGTTGGCTTGAAGCCAGCAGTACTCGGCGCTCTCGAGGAGTCCCTCTGCTC 3783
 DB 1 GGGCTGCTTGAAGCCCTGAGTGAATCTTGGCGGCTTCGCAAGGACGCGCTGTCTC 60

QY 3784 ATTGCCACCGCCAGACCAACCCGCTTTCATCACCAATAGCCAAAGAGGTACACAGACAT 3843
 DB 61 ATGCCACACGCCAGACCAACCCGCTTTCATCACCAATAGCTAAGGAGGTCCACAGACAC 120
 QY 3844 ACGGCTCTTGCAGCAAAATACCAATACAGCAGCAATATGACACACAACTCTTGCACGA 3903
 DB 121 ACGGCTCTTGCAGCAAAACCCAGTCCCAGCAGAGATATACACACCAACCCCTGGCGAGG 180
 QY 3904 GCTAAAGGGGAAATTTTGAGAGTCAATTGAAATCTTTATGAAAGATGCCACAGATGTT 3963
 DB 181 GCTAAAGGGGAAATCTGAGAGTCAATTGAAATCTCTATTGAAAGATGCCACGAGATGTG 240
 QY 3964 GTGATCTTCTCTGAGGATTTATGACATCATTTATGTACTGCTTGAAGGATCTTTAGTT 4023
 DB 241 GTGATCTTCTCTGAGGATTTATGATATCATCTACTGCTCTGGAAGGATCTTTAGTG 300
 QY 4024 AAAAGAAAGGCTTTCAGAAATGTTTTCCAGCCATCTGAGGTTCTACATGTTGTCAGTAT 4083
 DB 301 AAAAGAAAGGCTTTCAGGAGTGTTTTCCAGCTATCTGCAAGTTCTACATGTTGTCAGTAT 360
 QY 4084 TATGAGCGGAATCACAGAAATAGCAGTTGGAGCTCGCATGTTTTCAGTGGCCCTGTACGAC 4143
 DB 361 TATGAGCGGAGTCAAGAAATCGAGTTGGAGCAGCGCATGGCTCAGTAGCCCTGTATGAC 420
 QY 4144 ATCCGAGCTGGAATAATGTGACAAATTCATGGACACAGAGGACCAATCACTGAGTGCGT 4203
 DB 421 ATCCGAGCTGGAATAATGTGACAAATTCATGGGACACAGAGGACCAATCACTGAGTGTC 480
 QY 4204 TTGCTCTGTATGGAAGATATCTTGCCACTTACTTAAACACTGACAGCCACATTTCTTTT 4263
 DB 481 TTGCTCTGTATGAGCGGTTTACCTTGGCACCCTACTCGAACAACCGACAGCCACATTTCTTTC 540
 QY 4264 TGGCAGATGAACACGTCCTGCTGGGAAGCATCGGCATGCTGAATCTCGGACCTCAGCTG 4323
 DB 541 TGGCAGATGAACACCTCACTCTCTGGGAAGCATTTGGAATGCTGAATCTCAGCTCTCAGCTG 600
 QY 4324 CGCTGCATTAATAACCTTACAGAGTGCCCTCTGTGACGCGCGCTCCCGCTCCCAACAAT 4383
 DB 601 CGCTGCATCAAGACCTTACAGAGTGCCCTCTGAGTGACGAGCATCCCTGGCTCCCAACAAC 660
 QY 4384 GGCCTCAAGTGGCCGGCTCATCTGAGACTTCCAAACCGCAACGTCTATCTCATGCGCCAT 4443
 DB 661 GGCCTTAAGTTGGCCGGCTCATCTGAGCTTCCAAACCGGAATGTTATCTCTCATGAGCCAT 720
 QY 4444 GACGGGAAGGAGCAGCGCTTCTATGCTCTAA 4473
 DB 721 GATGGGAGGAACACCGCTTCTATGCTGA 750

RESULT 13

BX955125
 LOCUS
 DEFINITION DKFp781L11961_r1 781 (synonym: hlcc4) Homo sapiens cDNA clone
 BX955125
 ACCESSION DKFp781L11961 5', mRNA sequence.
 VERSION BX955125.1 GI:43437124
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 603)
 Wambutt, R., Heubner, D., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,
 Fobo, G., Han, M. and Wiemann, S.
 EST (Wambutt, R., Heubner, D., Mewes, H.W., Weil, B., Amid, C., et al.)
 Unpublished (2003)
 CONTACT: MIPS
 MIPS
 Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing Consortium of the German Genome Project.

No s1 sequence available.
This clone (DKFZp781l1961) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

source
1. .603
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp781l1961"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="781 (synonym: hlcc4)"
/notes="Vector: pSport1_Sfi; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN

Query Match 13.4%; Score 601.4; DB 5; Length 603;
Best Local Similarity 99.8%; Pred. No. 3.7e-162;
Matches 602; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2385 AACTATGGACACGTGCAAGCTGTTATGCTCCTTTCAGCGCTGGGGTTTGAATGAAGT 2444
DB 1 AACTATGGACACGTGCAAGCTGTTATGCTCCTTTCAGCGCTGGGGTTTGAATGAAGT 60
QY 2445 ACTGGATGAAGTTGCTGGATCGCTTGAATGCTGAAACCCCACTGCACCGTATCGTT 2504
DB 61 ACTGGATGAAGTTGCTGGATCGCTTGAATGCTGAAACCCCACTGCACCGTATCGTT 120
QY 2505 TGGCTCTTGTCAAGAGGAGCCATATGTCACGTATGCTGCGGGTTAATCAGCTGTC 2564
DB 121 TGGCTCTTGTCAAGAGGAGCCATATGTCACGTATGCTGCGGGTTAATCAGCTGTC 180
QY 2565 TTGTAAACTCTCATGTGGGAAACAGAAAGTAGGAAGAGCTGCCAGCGTCTCAGGGAGT 2624
DB 181 TTGTAAACTCTCATGTGGGAAACAGAAAGTAGGAAGAGCTGCCAGCGTCTCAGGGAGT 240
QY 2625 AGAAAGGGAACCTTACGAGTGTCCGTCGCTGCAACAGCATCTCTGCTATCAT 2684
DB 241 AGAAAGGGAACCTTACGAGTGTCCGTCGCTGCAACAGCATCTCTGCTATCAT 300
QY 2685 TTCTTTGGCAAACTTTATAGTATGACCAATGCACTTTTATTTGTTGATCATATGAA 2744
DB 301 TTCTTTGGCAAACTTTATAGTATGACCAATGCACTTTTATTTGTTGATCATATGAA 360
QY 2745 GAAGGGTCTTACAGGCACTAGACCAAGCACCAGACCTTTCTAAGCAAGGGGTTTC 2804
DB 361 GAAGGGTCTTACAGGCACTAGACCAAGCACCAGACCTTTCTAAGCAAGGGGTTTC 420
QY 2805 CCCTCCAACTTCCAGTAATATTGCAAGGACAGATTAACAAGTTGCTGCACTGTGCT 2864
DB 421 CCCTCCAACTTCCAGTAATATTGCAAGGACAGATTAACAAGTTGCTGCACTGTGCT 480
QY 2865 TTCCGCTCGGTCTGATGCTGATCACTTGCTCTGACCTCTCTGCTCTCTCTTACA 2924
DB 481 TTCCGCTCGGTCTGATGCTGATCACTTGCTCTGACCTCTCTGCTCTCTCTTACA 540
QY 2925 TACTCTGTTCTTAGTAATGAAGTTGGAGTGCAGTGTAGCTGTATGCACGTGTATGCT 2984
DB 541 TACTCTGTTCTTAGTAATGAAGTTGGAGTGCAGTGTAGCTGTATGCACGTGTATGCT 600
QY 2985 GCC 2987
DB 601 GCC 603

RESULT 14
CR544624
LOCUS
DEFINITION DKFZp459G1824_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone
631 bp mRNA linear EST 07-JUL-2004
DKFZp459G1824_5', mRNA sequence.

ACCESSION

CR544624

CR544624.1 GI:49896285

VERSION

ESI.

KEYWORDS

Pongo pygmaeus (orangutan)

SOURCE

Pongo pygmaeus

ORGANISM

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pongo.

REFERENCE

1. (bases 1 to 631)

AUTHORS

Koehrer, K., Beyer, A., Mewes, H.W., Weil, B., Amid, C., Oanger, A.,

Fobo, G., Han, M., and Wiemann, S.

TITLE

Pongo pygmaeus mRNA (Koehrer, K., Beyer, A., Mewes, H.W., et al.)

JOURNAL

Unpublished (2004)

COMMENT

Contact: MIPS

MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

This is the 5' sequence of the clone insert from S. Wiemann,

Molecular Genome Analysis, German Cancer Research Center (DKFZ);

Email s.wiemann@dkfz-heidelberg.de; sequenced by BMF (Biomedical

Research Center at the Heinrich-Heine-University,

Duesseldorf/Germany) within the cDNA sequencing consortium of the

German Genome Project. This clone (DKFZp459G1824) is available at

the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum,

Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email:

clone@rzpd.de Further information about the clone and the

sequencing project is available at

http://mips.gsf.de/projects/cdna/.

FEATURES

Location/Qualifiers

1. .631

/organism="Pongo pygmaeus"

/mol_type="mRNA"

/db_xref="taxon:9600"

/clone="DKFZp459G1824"

/tissue_type="cortex"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="459 (synonym: pcor1)"

/note="Vector: pSport1_Sfi; Site_1: SfiIA; Site_2: SfiIB"

ORIGIN

Query Match 13.4%; Score 601.2; DB 7; Length 631;
Best Local Similarity 97.8%; Pred. No. 4.3e-162;
Matches 609; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 3400 GGAGCTGAATTTGGTCTGAAATTTGAACCTCTTAACTATTGACACAGACTCGAAGCTCT 3459
DB 9 GGCNTGAAATTTGGTCTGAAATTTGAACCTCTTAACTATTGACACAGACTCGAAGCTCT 68
QY 3460 AGCCAAATTCCTGAGGAGATTCGGGTTGACTAGTGGTGGATCCAACTACTCGCTGGCCAGA 3519
DB 69 AGCCAAATTCCTGAGGAGATTCGGGTTGACTAGTGGTGGATCCAACTACTCGCTGGCCAGA 128
QY 3520 CATACTTGCAGGACTGAGTTTCTTCTGCTACAGCTTCCAGCCCAAACTTCTCTCCA 3579
DB 129 CATACTTGCAGGACTGAGTTTCTTCTGCTGAGCGGCTTCCAGCCCAAACTTCTCTCCA 188
QY 3580 CACAGCACTATCCGAAGAACAGCATTGATCTGATTGGACGTGGGTTCACTGTTGGGAG 3639
DB 189 CACAGCACTATCCGAAGAACAGCATTGATCTGATTGGACGTGGGTTCACTGTTGGGAG 248
QY 3640 CCTTACATGGATGTTCGCTGTTTCTGATGGGCTTCTCGAACTTTGTCCGATGCCGAG 3699
DB 249 CCTTACATGGATGTTCGCTGTTTCTGATGGGCTTCTCGAACTTTGTCCGATGCCGAG 308
QY 3700 AAACAATTCGCCAATCATCAATGGGTTGCTCTGAGCCAGCAGCTGACTCGGCCCGC 3759
DB 309 AAACAATTCGCCAATCATCAATGGGTTGCTCTGAGCCAGCAGCTGACTCGGCCCGC 368
QY 3760 TCTGCGAGGATGCGCTCTCGCTCATTTGCCCGCCAGACACCCCGCTTTCATCACCACC 3819
DB 369 TCTGCGAGGATGCGCTCTCGCTCATTTGCCCGCCAGACACCCCGCTTTCATCACCACC 428
QY 3820 ATAGCCAAAGAGGTACACAGACATACGGCTCTTTGAGGCAAAATCCCAATCAGCAGGAAT 3879

THIS PAGE BLANK (USP 10)